THE UNIVERSITY *of* EDINBURGH

Individual Tree Detection and Modelling Aboveground Biomass and Forest Parameters Using Discrete Return Airborne LiDAR Data

Wan Shafrina Wan Mohd Jaafar

Thesis submitted in fulfilment of

the requirements for the degree of

Doctor of Philosophy

to the

The University of Edinburgh

2017

Declaration

I declare that the work contained in this thesis is my own and has not been submitted for any other degree or professional qualification, except where explicitly stated otherwise in the text. In particular, research detailed in Chapter 4, 5 and 6 was adapted from papers that are either published or currently in review. I am lead author of the papers and the specific contribution from all co-authors are declared at the beginning of these chapters.

Wan Shafrina Wan Mohd Jaafar

October 2017

Dedicated to my husband,

Aizi Syam

To my three little musketeers,

Aaron, Aariz & Aarif

and my parents Rahdzah & Wan Mohd Jaafar,

Abstract

Individual tree detection and modelling forest parameters using Airborne Laser Scanner data (Light Detection and Ranging (LiDAR) is becoming increasingly important for the monitoring and sustainable management of forests. Remote sensing has been a useful tool for individual tree analysis in the past decade, although inadequate spatial resolution from satellites means that only airborne systems have sufficient spatial resolution to conduct individual tree analysis. Moreover, recent advances in airborne LiDAR now provide high horizontal resolution as well as information in the vertical dimension. However, it is challenging to fully exploit and utilize small-footprint LiDAR data for detailed tree analysis. Procedures for forest biomass quantification and forest attributes measurement using LiDAR data have improved at a rapid pace as more robust and sophisticated modelling used to improve the studies.

This thesis contains an evaluation of three approaches of utilizing LiDAR data for individual tree forest measurement. The first explores the relationship between LiDAR metrics and field reference to assess the correlation between LiDAR and field data at the individual-tree level. The intention was not to detect trees automatically, but to develop a LiDAR-AGB model based on trees that were mapped in the field so as to evaluate the relationships between LiDAR-type metrics under controlled conditions for the study sites, and field-derived AGB. A non-linear AGB model based on field data and LiDAR data was developed and LiDAR height percentile h80 and crown width measurement (CW) was found to best fit the data as evidenced by and Adj-R^2 value of 0.63, the root mean squared error of the model of 14.8% and analysis of the residuals. This paper provides the foundation for a predictive LiDAR-AGB model at tree level over two study sites, Pasoh Forest Reserve and FRIM Forest Reserve.

The second part of the thesis then takes this AGB-LiDAR relationship and combines it with individual tree crown delineation. This chapter shows the contribution of performing an automatic individual tree crown delineation over the wider forest areas. The individual tree crown delineation is composed of a five-step framework, which is unique in its automated determination of dominant crown sizes in a forest area and its

adaption of the LiDAR-AGB model developed for the purpose of validation the method. This framework correctly delineated 84% and 88% of the tree crowns in the two forest study areas which is mostly dominated with lowland dipterocarp trees.

Thirdly, parametric and non-parametric modelling approaches are proposed for modelling forest structural attributes. Selected modelling methods are compared for predicting 4 forest attributes, volume (V), basal area (BA), height (Ht) and aboveground biomass (AGB) at the species level. The AGB modelling in this paper is extracted using the LiDAR derived variables from the automated individual tree crown delineation, in contrast to the earlier AGB modelling where it is derived based on the trees that were mapped in the field. The selected non-parametric method included, k-nearest neighbour (k-NN) imputation methods: Most Similar Neighbour (MSN) and Gradient Nearest Neighbour (GNN), Random Forest (RF) and parametric approach: Ordinary Least Square (OLS) regression. To compare and evaluate these approaches a scaled root mean squared error (RMSE) between observed and predicted forest attribute sampled from both forest site was computed. The best method varied according to response variable and performance measure. OLS regression was to found to be the best performance method overall evidenced by RMSE after cross validation for BA (1.40 m^2), V (1.03 m^3), Ht (2.22 m) and AGB (96 Kg/tree) respectively, showed its applicability to wider conditions, while RF produced best overall results among the non-parametric methods tested.

This thesis concludes with a discussion of the potential of LiDAR data as an independent source of important forest inventory data source when combined with appropriate designed sample plots in the field, and with appropriate modelling tools.

Lay Summary

This thesis has investigated the novel techniques of airborne LiDAR based assessment for modelling forest tree attributes using the individual tree-based approach of a tropical forest in Peninsular Malaysia. The overall of the thesis is arranged as a procedure for AGB quantification as the main parameter of interest to derived from LiDAR using individual tree crown approach. Individual tree-based approaches are more prominent than area-based approaches, in terms of making the LiDAR inventory more closely resembling traditional field-based inventories but at much larger scale. The methods in this thesis contains three approaches on utilizing LiDAR data for individual tree measurement, from simple regression technique between LiDAR-derived height metrics and AGB at tree level based on trees that were mapped in the field (Chapter 4) to the methods of automated tree crown delineation (Chapter 5) and finally to the core modelling methods by using machine learning approaches for AGB and other important tree attributes which is Tree Height, Volume and Basal Area (Chapter 6). The success of this research has been centred to the novelty of LiDAR as an independent source with the aid of a complete set of field data and algorithm performance which has been proved to be effective to estimate important forest parameters, increased accuracy of the results and provided a scientific value for understanding the benefits of high density LiDAR data in individual tree crown (ITC) analysis for forest inventories. ITC and individual tree attributes are arguable of greatest interest to foresters as these are an important element that helps to describe the forest function, the productivity of forest ecosystem and optimized carbon sequestration.

Acknowledgements

I owe my deepest gratitude to the vast number of people and organisations who made it possible for me to complete this PhD. These include especially: Prof. Iain H. Woodhouse, Dr. Hamdan Omar and his team, Prof Zulkiflee Abd. Latif, Carlos Alberto Silva, Dr. Andrew Thomas Hudak, Prof. Johannes Breidenbach and Dr. Edward Mitchard.

Firstly, I would like to thank my principal supervisor Prof. Iain H. Woodhouse for his endless encouragement, support, advice, mentorship and faith in me over the past years. I never remember myself feel pressure or stress when I'm under your supervision during my PhD period. This is the first time I'm working with the most relax and easy-going person on earth!

Many thanks to Dr. Hamdan Omar from Forest Research Institute Malaysia for helping me during the fieldwork, kindly providing me with the best advice on field data collection and access to field datasets and LiDAR data covering the main study site at Pasoh Forest Reserve and second study site at FRIM Forest Reserve. My greatest appreciation when you bring your best man who expert in each area for assisting me in the collection of the entire field dataset in the challenging tropical rainforest environment at absolutely no cost. Mr. Azhan, Mr. Abu, Mr. Bohari, Yasmin and Mr Wan, which their name requires an acknowledgement here. I will be forever in your debt.

Special thanks to Prof Zulkiflee from Mara University of Technology, Malaysia who be the person suggesting me in choosing Pasoh as the study site for my PhD. Thank you for giving me the chance to be part of the NERC ARSF Flight campaign to Malaysia and the experience I received to witness the flight technical operations and sensors myself. Such a priceless experience. Thank you as well for the access to the fieldwork and survey tools through your university department.

Particular thanks to Carlos Alberto Silva, PhD student at University of Idaho, USA cum my research PhD partner in LiDAR who be the person that providing me with all the technical assistance in LiDAR data processing and always helping me to correct my R coding as well as always encourage me to publish paper and providing comments to my research article submitted to JTFS and Forest Journal.

To Dr. Andrew Thomas Hudak from US Forest Service (USDA) and Prof Johannes Breidenbach from Norwegian Forest and Landscape Institue go particular thanks for always patiently replying to my emails and providing comments to my research article and manuscript.

To Dr Edward Mitchard, thank you for your assistance during my first year of PhD especially on your advice on the fieldwork setting and providing me a hands-on tool on the use of fieldwork tools and giving me your best advice to help me shaping my research questions.

In my Edinburgh life, thanks are due to the past PhD students in Drummond Street, to the inhabitants of the Harris Suite and other to student in different school. You have been great companions and making me enjoy coming to office almost every day. Especially to Luise Fischer, David Rudolph, David Milodowski, Cha Younghwa, Emily, Vanessa Burton, Lucas, Siti and Noorzalianee. Some of you have leave a significant contribution such as providing me an assistance to proof read my manuscript, sharing programming code, technical run on super computer, words of support and encouragement. Special acknowledgement to Mawaddah, final year undergraduate student in software engineering, for your willingness to spend the whole day for a few occasions helping me to run Matlab and C++ code.

Finally, I would not be where I am now without the unwavering love and support of my family, in particular that of my husband, Aizi Syam. Your love, your support, your patience, your sacrifice, your care and concern are a motivation for me to keep going and finish this PhD. Special tribute to my angel of my heart, Aaron and Aariz who witness my struggle throughout this almost four years of PhD journey, put up with my long hours at school but still never missed to greet me with a hugged and kissed every time I came home and to the new addition to the family, Aarif, who was born a couple of months before completion of this thesis, you really motivate me to finish my writing quickly so that I can spend more time with you.

A heartful thank you to my wonderful parents (Rahdzah and Wan Mohd Jaafar) for always keeping me on their prayers and having faith in my ability to finish this thing even when I didn’t. Thank you to the love and supports from my sister, Molyna and brothers, Hafidz and Haziq and thank you to my parent in law (Aminah and the late Ibrahim) and to all my family in law members for your endless support. Last but not least, to Athirah whom I treat as my own little sister, thank you for always be our saviour, looking after my kids and love them like your own family.

Finally, many thanks to all the other people not listed here who have assisted me in producing this thesis: I am sorry there is no room to list you all here, but your help is greatly appreciated.

Table of Contents

CHAPTER 3
The Study Area: Pasoh Forest Reserve and FRIM Forest Reserve

CHAPTER 4
Modelling Individual Tree Aboveground Biomass Using Discrete Return LiDAR in Lowland Dipterocarp Forest of Malaysia

CHAPTER 1

Introduction

1.1 Aim

The aim of this thesis is to investigate the use and effectiveness of novel airborne LiDAR (Light Detection and Ranging) technologies to extract metrics from individual trees and to model aboveground biomass (AGB) and forest parameters in tropical rainforests of Malaysia. Forests constitute a major part of terrestrial ecosystems, occupying approximately 30% of the world's land area (FAO 2010). Forests absorb carbon through photosynthesis and sequester it as biomass, thus contributing significantly to the mitigation of global climate change, and it is estimated that over 80% of the world's aboveground and 40% of belowground carbon are stored in forest ecosystem (Dixon *et al.* 1994). Because of this, the United Nations Framework on Climate Change (UNFCCC) has created a collaborative program on Reducing Emissions from Deforestation and forest Degradation (REDD+) in developing countries, in order to establish efficient systems for monitoring forest carbon stocks (UNFCCC 2007). Malaysia's interest in participating in the World Bank's Forest Carbon Partnership Facility (FCPF) were required to prepare a REDD Project Idea Note (R-PIN) for the preparation of a detailed Readiness Preparation Proposal (R-PP) as a baseline study to present country's current status of carbon stored in forest (Hamdan 2012). The development of a biomass model for carbon stock estimation will

help to support the research development for carbon monitoring methodology in Malaysia with the ultimate of aim of preparing REDD+ implementation in Malaysia (FRIM 2011, FRIM 2012). The increasing importance of accurate biomass estimation to support the REDD+ implementation, has created a critical need to understand, evaluate and improve current tree biomass prediction methods.

Remote sensing has the potential to provide quantitative information for estimating biomass and biomass components (Hollaus *et al.* 2009), and has been shown to be a cost-effective tool to provide temporally uniform carbon stock observations over time (Wulder *et al.* 2010). Remote sensing for forest biomass estimation has traditionally focused on the use of imagery from passive optical and RADAR (Radio Detection and Ranging) sensors in the tropical forests (Clark *et al.* 2011; Du *et al.* 2014). However, there is a limitation of these sensors in terms of sensitivity, accuracy, environmental conditions and information on the vertical structure of the forests. LiDAR has the capability to provide detailed, spatially explicit, three-dimensional information on forest canopy structure is highlighted to play a significant role to overcome the saturation problem of other sensors in high-biomass forest estimation.

LiDAR laser pulses that can penetrate through even the dense multi-layered canopies and retrieve vertical structure information from return signals results to be a strong correlation between LiDAR data and many forest biophysical properties (Lu *et al.* 2012). The potential of LiDAR to accurately estimate forest biomass and other important forest attributes have been demonstrated in boreal, temperate and tropical forest and in airborne LIDAR data, metrics can be extracted on the basis of either individual trees or areas (Chen, 2013).

There is a particular need for assessing biomass and other tree biophysical parameters at tree level using LiDAR data. Such assessments are needed by forestry departments for the purpose of forest inventory and forest management implemented at plot-level scale. The Forest Research Institute of Malaysia (FRIM), for example, have developed a model at individual tree scale to access information for each tree, and on a tree species basis. Model developed by LiDAR at tree level could be of useful to extract important parameters such as tree height and crown size for more accurate biomass estimation. The advent of LiDAR imaging techniques coupled with advanced statistical techniques ranging from simple regression between LiDAR-derived height metrics and forest AGB to methods including automated tree crown delineation, stochastic simulation and machine learning approaches, have resulted in a number of different studies exploring their potential in deriving accurate biomass among other forest studies (Gleason & Lm 2012).

This thesis presents the results of research on the use of Airborne LiDAR to delineate individual trees and to model AGB and forest parameters such as height, basal area and volume in the tropical rainforests of Malaysia. This study initially focuses on the capability of discrete return LiDAR sensor to determine AGB at tree level based on trees that were mapped in the field. This method provides a good guideline for forest managers and forest inventory work at tree level, this is especially useful for developing countries like Malaysia, which may not have resources for extensive field campaigns of forest surveying. The later work scales up to automatic crown delineation and the implementation of an AGB model that was developed at the tree level to validate the results for the automated method. This chapter helps to answer on the context of remote sensing and to show the importance of correct delineation of tree

crowns on AGB estimation and on the last chapter, modelling AGB and other forest parameters by comparing the accuracies between parametric and non-parametric approaches. The results are arranged into three chapters that ultimately form three research papers. These papers, when taken together, assist with the process of developing a framework describing how the stages of modelling individual tree AGB based on the target users and the implementation.

The aims of this research can be summarised as:

1) To identify the relationship between AGB measured in the field and LiDAR metrics, and how these relationships impact the accuracy of predictive model at tree level;
2) To develop a framework methodology to improve the accuracies of individual tree crown delineation by taking advantage of both the simplicity of the canopy height model (CHM)–oriented methods and detailed 3D structures of tree crowns.
3) To compare the parametric approach and non-parametric approach in modelling AGB and other forest attributes parameters such as basal area, stem volume and tree height.

1.2 Why detect individual tree and modelling AGB and forest parameters over tropical forests using Airborne LiDAR?

1.2.1 The role of forests in the global carbon cycle

Vegetation, especially trees, removes carbon from the atmosphere and stores it as biomass. If the forest is cleared, or becomes degraded, the carbon can be released back

to the atmosphere, either through burning or decomposition. Anthropogenic carbon emissions to the atmosphere were 555 ± 85 PgC between 1750 and 2011. Of this amount, fossil fuel combustion and cement production contributed 375 ± 30 PgC and human-induced deforestation and forest degradation is estimated to be 180 ± 80 PgC (Ciais et al. 2014). Although the proportion of emissions stemming from land use and cover change has a decreasing trend (Cias et al. 2014), land use and land cover change is still a significant source of carbon emissions with an estimated total of 0.87 ± 0.49 PgC for 2013 (Friedlingstein et al. 2010).

Tropical forests, found around the equator between latitudes of 23ºN and 23ºS, cover around 18 million km^2 (FAO 2011) and are estimated to store 271 ± 19 PgC (Grace et al. 2014). These forests are under great pressure for conversion to agricultural land (Houghton, 2012), and Grace et al. (2014) report a total carbon loss of 2.01 ± 1.1 PgC yr^{-1} from deforestation, harvesting and peat fires. However, the growth in forests and woodlands is reported to sequester 1.85 ± 0.09 PgC yr^{-1} (Grace et al. 2009) resulting in a net loss of 0.16 ± 1.1 PgC yr^{-1} from tropical forests. Thus, despite being a substantial potential carbon sink, approaching 2 PgC yr^{-1} or up to 20% of the global carbon emissions, their net loss makes them a carbon source.

With the prospect of a quick and cheap solution for mitigating carbon emissions (Gibbs et al. 2010), tropical forests have received much attention and have resulted in the policy and economic incentive mechanism known as the REDD+ mechanism. REDD+ (reducing emissions from deforestation and forest degradation, for conservation and enhancement of forest carbon stocks), described in the 16th session of the Conference of parties to the United Nations Framework Convention on Climate Change, gives developing countries the opportunity to monetize the reduction of

emissions from deforestation and forest degradation, and enhancement of forest carbon stocks (UNFCCC, 2011).

Accessing carbon finances through REDD+ will require, among other factors, robust measurement of carbon stock changes in forests (UNFCCC, 2010). Furthermore, a mechanism for commercial trading of forest carbon credits earned through enhancement of forest carbon stocks, conservation of forests or sustainable forest management require trustworthy systems for verification of carbon offsets. In addition, the uncertainty of estimates to minimize the risk of overestimating emission reductions (UNFCCC, 2006) and lack of accurate biomass estimates may result in loss of carbon credits for the project developer (Gibbs et al. 2007).

1.2.2 Forest inventories supported by remotely sensed data

Forest inventories that are based on field sample surveys, supported by auxiliary remotely sensed data, have the potential to provide reliable and confident estimates of aboveground biomass, hereafter simply referred to as AGB. This method also provides more consistent and accurate estimates of other forests attributes e.g. stem volume and basal area.

Forest inventories are usually designed as sample surveys, with observations on the ground collected from field plots. Alternatively, aerial photography has been an important forest inventory tool since the 1940s and more recently the availability of optical satellite images since the 1970s has allowed global forest cover statistics (Boyd & Danson 2005). While high costs have prevented the use of aerial photography, the use of low-cost optical satellite images has been hampered by low spatial resolution

and persistent cloud cover in tropical areas. Furthermore, both aerial and satellite optical images have traditionally only provided two-dimensional information, although recent developments have resulted in three-dimensional (3D) data from aerial and satellite images with the use of digital photogrammetry and image matching (e.g. Næsset 2002; Bohlin et al. 2012; Persson et al. 2013; Gobakken et al. 2015).

The use of LiDAR (Light Detection and Ranging) sensors most commonly mounted on a small aircraft and with a scanning capability, known as airborne laser scanning (ALS), has proved to be effective for accurately determining AGB in different forest types (Zolkos et al, 2013; Fassnacht et al. 2014). There has been a strong focus on research of airborne LiDAR during the past two decades, and LiDAR is now used as an integral part of operational forest management inventories in several countries (McRoberts et al. 2010; Næsset, 2014). Most of the published studies on LiDAR to estimate biomass have been carried out in boreal and sub-boreal coniferous forest with relatively low biomass and open forest structure. However, in the last five years, use of LiDAR for biomass estimation has been demonstrated in tropical forests in South America (Vincent et al. 2012; Andersen et al. 2014; Asner & Mascaro 2014), Asia (Jubanski et al. 2013; Ioki et al. 2014) and Africa (Asner et al. 2012). The maximum biomass densities in these studies were about 500 metric tonnes of biomass per hectare ($Mgha^{-1}$), while biomass densities in tropical rainforests can reach levels beyond 500 $Mgha^{-1}$ (Keith et al. 2009).

1.2.3 Sources of three-dimensional remotely sensed data

Satellite-mounted optical sensors have been used for estimation of forest cover since the launch of the first Landsat satellite in 1972 (Iverson et al. 1989). More importantly, with data spanning over three decades, these optical sensors provide estimates of global forest cover change (Hansen et al. 2013). For biomass estimation in the tropical forests, estimation has traditionally focused on the use of imagery from passive optical and RADAR (Radio Detection and Ranging) (Clark et al. 2011; Du et al. 2014). The sensitivity and accuracy of these sensors have been shown to decrease with increasing aboveground biomass and leaf area index, which inhibits reliable carbon stock estimates (Tsui et al. 2012). Furthermore, the usefulness of the two-dimensional information from satellite images is limited because it lacks information on the vertical structure of the forests, has limited spatial resolution, and is often obstructed by cloud cover.

LiDAR systems solve these challenges by emitting a short pulse of laser light and measuring the time between the emission and the reflectance (echoes) detected by the LiDAR sensor. By emitting thousands of pulses per second and recording several echoes per pulse in scanning motion, the LiDAR system effectively creates a 3D cloud of echoes. By recording the position and orientation of the sensor at the time of emitting each pulse, using a GPS (Global Positioning System) receiver and an inertial navigation system unit, each echo is positioned in the 3D space (x, y, and z positions). To derive information about the vegetation, a digital terrain model (DTM) is constructed by identifying those echoes that come from the ground. Following the construction of the DTM, the elevation of all echoes in relation to the DTM is

computed. Echoes above a certain threshold height above the DTM (which is context specific) are regarded as vegetation echoes.

Although the best results for biomass estimation have been obtained using LiDAR, its cost is high compared to using satellite-based sensors (Böttcher et al. 2009). The cost will depend on the requirements within REDD which is mainly determined by the accuracy level. The accuracy level will determine the monitoring technology applied, each requiring different ways of data acquisition, processing, training and capacity building. The factors that influence the price are the amount of earth observation data needed. Table 1 lists costs of acquisition and analysis of monitoring services from various technologies of earth observation based on the most recent years of available information.

Table 1.1. Acquisition and analysis costs* of monitoring services of various earth observation technologies in US$ in 2009. Adapted from "An assessment of monitoring requirements and costs of Reduced Emissions from Deforestation and Degradation", by Hannes Böttcher, Katja Eisbrenner, Steffen Fritz, Georg Kindermann, Florian Kraxner, Ian McCallum and Michael Obersteiner. *Carbon Balance and Management*. 2009. Copyright BioMed Central.

Satellite and sensor	**Resolution and coverage or project area**	**Costs for data acquisition**	**Cost for analysis**	**Total monitoring costs**	**Source**
Optical, medium resolution sensors					
Landsat -5, TM	30 m, 180 x 180 km	0.02 US$/km^2 – free	Classification 0.12– 0.31 US$/km^2 Change detection 0.4–0.6 US$/ km^2	0.50–1.21 US$/ km^2	SARMAP pers. comm
Landsat-7, ETM+ SPOT 4	30 m, 60 × 180 km 20 m	0.06 US$/km^2 0.31 US$/km^2			

Terra ASTER	15 m, 60 × 60 km	0.02 US$/ km^2			
CBERS-2, HRCCD	20 m	free in Brazil			
DMC	32 m, 160 × 660 km	0.04 US$/ km^2			
IRS-P6-LISS III	23.5 m	0.07 US$/ km^2	Human resources and equipment 0.5 US$/ km^2	0.57 US$/ km^2	UNFCCC
Optical, high resolution sensors					
Quickbird	3 m	25 US$/ km^2	Classification 2.2–2.5 US$/ km^2 Change detection 4.7–7.9 US$/ km^2	7.50 – 35.40 US$/ km^2	SARMAP pers. comm.
Ikonos	4 m	25 US$/ km^2			
RapidEye	5 m	2.8 US$/ km^2			RapidEye pers. comm.
SPOT-5, HRVIR	5–20 m, 60 × 60 km	0.6 US$/ km^2			SARMAP pers. comm.
Optical, very high resolution sensors					
Quickbird	0.6 m	16–22 US$/ km^2	Classification 100–125 US$/ km^2	116–272 US$/ km^2	SARMAP pers. comm.
WorldView-1	0.5 m	16–22 US$/ km^2	Change detection 160–250 US$/ km^2	116–272 US$/ km^2	SARMAP pers. comm.
Radar, SAR					
ALOS PALSAR	10-15 m	0.04 US$/ km^2	Classification 2.2–2.5 US$/ km^2	6.94 – 10.44 US$/ km^2	SARMAP pers. comm.
Satellite or shuttle SAR		0.14 US$/ km^2	Change detection 4.7–7.9 US$/ km^2	7.04 – 10.54 US$/ km^2	Patenaude *et al.* 2005
Airborne SAR		345 US$/ km^2		>345 US$/ km^2	Patenaude *et al.* 2005

LiDAR, Airborne					
UK, forest monitoring, national average	28,000 km^2			415 US$/ km^2	Patenaude *et al.* 2005
US, forest inventory at project level	40 km^2			455 US$/ km^2	Tilley *et al.* 2004
	400 km^2			100 US$/ km^2	Tilley *et al.*2004
US project area	180 km^2			388 US$/ km^2	Parket & Evans 2007
Indonesia, forest inventory at project level	136 km^2	400–550 US$/ km^2	160 hours processing time	> 400–550 US$/ km^2	RSS GmbH pers. comm.

* Costs for analysis and total costs are indicative costs. This include service design, data processing and mapping, interpretation and analysis. The actual costs would depend on the selected sensor, the fit of sensor data to area to be mapped (which determines how many scenes are needed), the amount of GIS (Geographical Information System) processing, integration and support services required to develop final images and maps and integrate these into asset operational and management systems.

Satellite-based optical imagery is frequently obstructed by persistent cloud cover in the tropics, use of active synthetic aperture radar (SAR) sensors penetrate clouds and produce backscatter images that can be used for the prediction of forest biomass. In high biomass conditions, however, radar backscatter data is not able to provide reliable estimates of AGB, and has been shown to saturate at biomass levels between 200-250 $Mgha^{-1}$ even for very long wavelengths. (Mitchard et al. 2009; Le Toan et al. 2011). Promising results have nevertheless been published for biomass values up to 450 $Mgha^{-1}$ using long wavelength (P-band) SAR tomography, (Minh *et al.* 2014) which is expected to be achievable with the future BIOMASS Earth Explorer mission from ESA (European Space Agency) (Fayad et al. 2014). The saturation problem in carbon stock estimation in high-biomass forests can be overcome by LiDAR, as laser pulses can penetrate through even the dense multi-layered canopies and retrieve vertical structure information from return signals, and as a result there tends to be strong

correlation between LiDAR data and many forest biophysical properties (Lu et al. 2012).

At present, SAR technologies exist that can produce 3D data using four different techniques: clinometry, stereoscopy, interferometry and polarimetry (Toutin & Gray 2000). In addition, optical satellite images can produce 3D data by repeat-pass-acquisition and image matching techniques. New applications are being developed continuously and a thorough overview is beyond the scope of this thesis. A shared property of these techniques is that, in order to provide information at a level similar to that of LiDAR, they require a high-quality DTM. At present, the only technology able to provide this DTM quality is LiDAR, and it is therefore a prerequisite to optimise the use of the other sensors and techniques. Although P-band InSAR or tomography can retrieve a a ground surface, it is never as accurate as with LiDAR. Previous studies have demonstrated the potential of LiDAR to accurately estimate stand level forest biomass in boreal (Næsset & Gobakken 2008; Næsset et al. 2013; Kankare et al. 2013), temperate (Lefsky et al. 1999; Lim & Treitz 2004) and tropical forests (Drake et al. 2002; Asner et al. 2010), as well as biomass at the single tree level (Popescu, 2007; Haughlin *et al.* 2013).

1.2.4 Useful LiDAR Metrics for Biomass Estimation and Statistical Models

The most widely used LiDAR metrics for biomass prediction are various height metrics (Lim *et al.* 2003; Lim & Treitz 2004; Patenaude *et al.* 2004; Hall *et al.* 2005; Wan-Mohd-Jaafar et al. 2017). LiDAR metrics can be calculated based on first, last, or all returns (waveform metrics) although not one single metric can predict biomass

well over different sites. Unfortunately, the best height metric reported in the literature differs between each of the previous studies. These include the 80th (Patenaude *et al.* 2004; Wan-Mohd-Jaafar et al. 2017), 75th (St-Onge et al.2008), 50th (Thomas *et al.* 2006), 30th (Stephens *et al.* 2012), and 25th (Lim & Treitz 2004) percentile heights. The differences in vegetation structure, model development and data processing procedure may result such variability but the dominant cause is unclear. A study conducted by Lim & Treitz (2004), St-Onge et al. (2008) and Thomas et al. (2006), all focused on boreal forests, but their metrics for biomass estimation differ significantly, these results could be coming from the data processing procedure or model development selection. These three studies differ in model form and data processing procedure: Lim & Trietz (2004) and St-Onge et al. (2008) used a simple power model, whereas Thomas et al. (2006) used a simple linear model with square root—transformed biomass; Lim & Treitz (2004) and Thomas et al. (2006) derived height metrics from a raw point cloud, whereas St-Onge et al. (2008) calculated height metrics based on CHM (Canopy Height Model). In particular, the percentile height in the study by Lim and Treitz (2004) was calculated based on overstory canopy returns, which might explain why a much lower percentile was chosen. The best metrics in the studies of Patenaude *et al.* (2004), St-Onge *et al.* (2008) and Wan-Mohd-Jaafar et al. (2017) are similar (80th and 75th percentile heights, respectively) but the vegetation structure of the study site (focused on temperate deciduous forest, boreal forest and tropical rainforest respectively) and the data processing procedure is different where percentile heights were extracted from a 1×1 m^2 CHM in a studies by Patenaude *et al.* 2004 and St-Onge *et al.* 2008 while Wan-Mohd-Jaafar et al. (2017) derived height metrics from the raw point cloud data but the biomass was modelled with a similar

power model. The sites in this study are both from a very complex environment and vegetation structure. Assessment on which LiDAR metrics correlate best with biomass at tree level is crucial, especially for forest inventory and forest manager for the purpose of forest management. Height metrics also has been successfully used for predicting stem volume and basal area (Chen et al. 2007).

Power models have been used in a large number of studies for biomass estimation, probably because most allometric equations for calculating biomass in the field are power models. At the log–log scale, power models (Lim et al. 2003; Lim & Treitz 2004; Patenaude et al. 2004; St-Onge et al. 2008) correspond to linear regression models, whereas multiplicative power models (Naesset & Gobakken 2008) correspond to multiple linear regression models. Many studies have performed nonlinear transformation on LiDAR metrics for biomass. For example, the quadratic term (squared term) of LiDAR metrics might be used (Asner 2009). Biomass might be log transformed (Hall et al. 2005) or square root transformed (Thomas et al. 2006) for reducing the heterogeneity of regression residual variance. Although nonlinear models have been widely used, there are a few studies that have used linear models for predicting biomass (Bortolot 2006; Lucas et al. 2006; van Aardt et al. 2006; Asner et al. 2009; Zhao et al. 2009).

1.2.5 Individual tree-based method for forest inventories and biomass estimates using LiDAR data.

LiDAR data is an important complement to other remote sensing sources for information extraction at individual tree level (Koukoulas & Blackburn 2005;

Magnusson 2006; Maltamo et al. 2006). In a diverse forest, such as those found in the tropics, a stand-wise approach is usually not sufficient for forest management planning, such as the established methods in a number of European countries (Koch et al. 2006). This is due to the different growing behaviour of several tree species can occur in one stand, a priori knowledge of stem number and tree species distribution would be necessary for calculating stand parameters. The tropical forest planning systems typically work at the single tree level (e.g. Lamas & Eriksson 2003) and in harvest management, information of single tree is required. Therefore, single tree detection for the purpose of estimation of single tree biomass and related information extraction is a prerequisite to fulfil these needs.

Discrete return and waveform LiDAR have been widely applied for forest height, crown volume and biomass estimation. While medium or large footprint (20-70 m) LiDAR data are useful for characterizing the vertical distribution of canopies at the resolution of the footprint, small footprint (10s of cm) LiDAR provides both vertical and horizontal information at the scale of individual trees (Wulder et al. 2012). Estimates of forest biomass have largely ignored the highly detailed spatial information from discrete return LiDAR and focused on metrics such as canopy height and cumulative vertical distributions at plot level, as in Chen et al. 2007; Popescu et al. 2003; and Popescu (2007). Providing more spatially detailed information such as the number, location, spacing and size distribution of individual trees may improve biomass estimation at varying spatial resolutions, and should provide a more ecologically meaningful structural description of a forest.

The crown is the basic meaningful object in any application and analysis at the individual tree level. Various method for extracting individual tree information from high resolution LiDAR datasets have been developed. Through time, studies on individual tree detection have shown increased complexity of analyses, increased accuracy of results and a focus on the use of LiDAR data alone.

These techniques generally fall into three categories; local maxima detection and expansion (Kaartinen et al. 2012; Leckie et al. 2003; Maltamo et al. 2004; Persson et al. 2002; Popescu & Wynne 2004; Vastaranta et al. 2011; Wulder et al. 2000), watershed-based delineation (Breidenbach et al. 2010; Chen et al. 2006; Koch et al. 2003; Kwak et al. 2007), and point cloud based (Rahman & Gorte 2009; Li et al. 2012; Ferraz et al. 2012; Ferraz et al. 2016).

Local maxima algorithms typically involve the selection of search radius and detection of local maxima from a Canopy Height Model (CHM). Popescu & Wynne (2004) used both circular and square windows with site-specific window sizes to increase local accuracy of maxima detection. Leckie et al. (2003) applied a valley-following approach to isolate crowns based on CHM topography that yielded both tree locations and crown geometries with 80% accuracy. Vastaranta et al. (2011) used a minimum curvature approach with local maxima detection for a boreal forest and although they did not present an individual tree accuracy, they used delineated crowns to predict basal area ($R^2 = 0.48$) and volume ($R^2 = 0.71$). Maltamo et al. (2004) also worked in a boreal forest with a local maxima detection algorithm and reported that while as much as 80% of dominant crowns were correctly detected, the total accuracy was 40% due to issues identifying understory crowns. Although local maxima techniques are computationally the fastest and simplest algorithms, these algorithms have difficulty

detecting crown edges, typically oversimplifying crown geometry (Kaartinen *et al.* 2012).

Watershed-based delineations offer an improvement for crown geometries, and CHM image grey tone is inverted so the local maxima become local minima and vice versa (Chen et al. 2006). Some issues remain to be resolved especially for delineation of trees with complex structures. For instance, over-segmentation can be occurred due to the branches and sub-crowns of a tropical trees may resemble small trees. Tropical trees are close and overlapping each other makes between-crown valleys so invisible that a tree cluster could be falsely detected as one crown leading to under-segmentation.

To improve the individual tree crown delineation, attempts to segment individual trees using LiDAR 3D point clouds have been conducted (Reitberger et al. 2009; Li et al. 2012; Ferraz et al. 2012; Ferraz et al. 2016; Hu et al. 2017). Point cloud based techniques are the newest and most computationally demanding of the tree delineation approaches. The detailed methods are described in Chapter 2. Because of point cloud based methods utilized the detailed 3D point cloud information directly, these methods working on a massive data points and require high computation power, and it is very challenging to extract useful crown features from LiDAR returns generated by various objects in the forest (Li et al. 2013).

Most of the approaches discussed have different drawbacks. This thesis (Chapter 5), addresses the development of methods to improve the accuracies of individual tree crown delineation by take advantage both the simplicity of the CHM-oriented methods and detailed 3D structures of tree crowns. Watershed segmentation is prone to under

or over-segmentation due to difference in tree heights and natural variability of vegetation within tree crowns, hence, marker-controlled watershed segmentation routine used to overcome this problem. The basic idea in this approach is to mark the trees and guide the watershed procedure to only delineate those marked trees, but to mark manually is impractical for a large-scale data. Mathematical morphology was introduced as an automated approach to mark the tree apexes. To improve the efficiency of the watershed results, a detailed examination of the 3D LiDAR points was examined only to selected segments that was identified as "problematic" based on prior knowledge of the study area and by applying a few rules and parameter to detect the trees that need to be segmented again.

1.2.6 Improving the accuracy of biomass estimates and the need for better estimation approach

As described in section 1.2.1 accurate biomass estimates are a requirement for effective forest management, and for evidence to support the REDD+ mechanism. Increased accuracy would also potentially lead to added carbon credits for the project developer (Gibbs *et al.* 2007). Accuracy is defined as the sum of "trueness and precision" (International Organization for Standardization [ISO], 2012). Accuracy of an estimation is often expressed by the mean square error (Gregoire & Valentine, 2008), or the root mean square error (RMSE) of the mean estimate. Thus, accuracy incorporates both trueness, expressed herein by the mean difference (MD), and precision, expressed as standard error of estimation (SE), i.e., the square root of the estimation variance, or standard deviation of a sample (SD). The simplest way of

increasing the precision of biomass estimates is by increasing the sample size. Doubling the number of observations would halve the variance of the estimation. In this thesis, the study addresses the issue by assessing the robustness of biomass estimation models based on LiDAR data using models calibrated with similar field sample data collected at two different sites that are similar in terms of environment, geographical factor and forest structure. Pooling the sample data from both site could reduce the issue of having small sample data to build a reliable regression model to represent all species in the dense tropics.

1.2.7 Comparing modelling methods for predicting Forest biomass and forest parameters estimates

FRIM and Forestry department of peninsular Malaysia has been looking into a cost-effective and accurate forest inventory data and this has led to new ways of estimating and imputing plot data collected by Malaysia NFI (National Forest Inventory). Forest managers need accurate forest inventory data to develop a forest management plan that will allow them to prepare for future forest activities where often, these data must cover up to thousands of acres of land areas and the cost to collect them can be very high. Without an accurate estimation method, land managers might end up with incorrect biomass estimate maps, which could lead them to make poorer decisions in their future management plans (Gagliasso et al. 2014). As to meet national and international negotiations and reporting requirements, forest management plans require local inventory data on biomass, carbon and other resources. The data must be intensive enough to include structural variables relevant to biomass and extensive

enough to cover hundreds to thousands of hectares at a very minimal cost (Gagliasso et al. 2014). Recognition of the widespread need for cost-effective, local inventory data that spans large region has led to new methods for imputing plot data to site without data and then generating estimation of regional biomass. Parametric and non-parametric methods can be accessed to predict forest attributes that is useful for forest biomass estimation. Methods including Ordinary Least Square (OLS) regression, Gradient Nearest Neighbour (GNN) imputation, Most Similar Neighbour (MSN) imputation and Random Forest (RF) are one of the most widely used method to predict biomass and basal area in many studies (Temesgen & Hoef 2015; Mauya et al. 2015), attempt of this approach considered as one of the first approach in the region.

1.3 Overview of the thesis

This chapter has presented an introduction to the subject area and the urgency of this research, as well as an overview of the structure of the thesis. This chapter relates the AGB estimation in the context of international issues and outlines the aims. The remote sensing approaches and technical context used in this thesis to estimate AGB using LiDAR are elaborated in Chapter 2. This followed by the description of study area in which the location, characteristics and condition of the forest are discussed in Chapter 3. *In-situ* methods for the AGB estimation are also discussed as the fieldwork data are used to validate the result of AGB estimation derived by LiDAR. This chapter also outlines the specification of the LiDAR data used in this research.

The next three chapters are each based on stand-alone papers, which have either been published or are at the time of writing going through the peer-review process. Chapter 4 defines the methodology of AGB estimation using LiDAR and integration with the field sample data as direct approach, to see the relationship and to derive the most significant LiDAR metrics that correlate well with AGB to develop the predictive LiDAR model. The results are discussed accordingly in the chapter.

Chapter 5 presents the approach for automatic extraction of individual tree crown delineation by integrating the CHM-oriented approach and point-cloud based approach as for refinement of the results. The accuracy of the results was validated by using the model developed in previous model and with the ground reference data and discussed at the end of the chapter. Chapter 6 discusses in more detail the derivation of AGB and tree biophysical parameters by integrating full modelling and comparison are made between parametric and non-parametric approaches. Finally, Chapter 7 synthesises the findings of the preceding chapters and presents a vision of using LiDAR data for extracting individual tree biomass and forest biophysical properties at variety of methods and accuracy. This chapter concludes the thesis with recommendations for future work.

CHAPTER 2

Introduction to Discrete Return Airborne LiDAR data for Individual Tree Detection and Modelling Aboveground Biomass and Forest Parameters of a Tropical Rainforest

2.1 Introduction

This chapter provides a general introduction to the research field for non-specialist, and introduces and explains many of the concepts and terms used in later chapters. In particular, it provides an overview of the characteristics of tree parameters measured in the field, an introduction to the Airborne LiDAR data used as a main tool to extract individual trees and the technical approach in modelling aboveground biomass and important forest parameters using LiDAR metrics and ground measurements. From the overview, a full range of research gaps has been highlighted which have been used to formulate research aims (Section 1.1).

2.2 Characteristics of Forest Parameters

Before discussing the technical details of Airborne LiDAR systems and the methodology used to estimate the forest parameters of interest, it is important to correctly define the ecological variables that the author interested in, and how they are measured on the ground. Most of the chapters in this thesis concerned with aboveground biomass (AGB) but Chapter 6 discusses on modelling other important

forest parameters that are also relevant for the forest studies. Here these concepts, and the techniques used to measure them in the field, are defined.

2.2.1 Tree height and bole height

Tree height is defined as the distance from the ground to the topmost leaf. It can be very difficult to measure, especially in a dense tropical forest, but as long as the top and bottom of the tree is visible it can be determined relatively accurately with a vertex hypsometer or clinometer (Phillips et al. 2009).

Tree height is largely used in this thesis in combination with Diameter at Breast Height (DBH) and wood density in order to estimate the AGB of a tree. However, it can also be used as a parameter in its own right to characterise a forest. Two metrics are often used to summarise tree height data within a plot; these are:

a) Total Height – the vertical distance from the base of the tree to the uppermost point (tip).
b) Bole Height – the distance between tree base to the first live branch on the main stem.

These two metrics can be estimated with relative accuracy from LiDAR data given sufficient sampling density. Tree height are usually well correlated to AGB within a particular landscape, this provides a useful means to produce estimate of AGB from LiDAR data.

2.2.2 Basal Area

Basal area is often used to compare different wooded areas, especially by forest managers and ecologist. In a certain forest types, basal area tends to correlate well with AGB and it is defined as the area of a vertically-projected forest that would be tree-stem if the forest was sectioned at 1.3 m above the ground. Normally only stems with a diameter greater than 10 cm are considered, as in most forests stems smaller than 10 cm represent only a small fraction of the basal area (Chave et al. 2003). It can be easily calculated by using the standard primary measurement collected in field forestry that is from the measurement of DBH (diameter at 1.3 m). From this, basal area per tree can be calculated using the Equation 1, where basal area is *Ba* (m^2) and DBH (cm) (Loetsch et al. 1973):

$$Ba = \frac{DBH^2 * \pi}{40000}. \quad (1)$$

The basal area per ha (m^2/ha) was obtained by the sum of individual basal area, A (ha), corrected for the size of each individual plot (Equation 2):

$$BA = \frac{\sum Ba}{A}. \quad (2)$$

Basal area estimates can be biased easily by incorrect DBH measurement methodologies. For example, not correctly treating multiple stems, or measuring over buttresses and other stems irregularities, can cause large errors. For this reason, strict guidelines should be followed, for example those laid out in the RAINFOR field manual (Phillips et al. 2009).

2.2.3 Aboveground Biomass (AGB)

AGB is the principle parameter of interest throughout this thesis, and it is defined as the mass of aboveground living material per unit area, expressed in tonnes biomass per

hectare (Mg/ha). AGB is very time-consuming and destructive to measure directly e.g. measuring AGB for a one ha plot would involve cutting down and weighing at least several hundred trees. In more practical approach, AGB is normally estimated by measuring a few parameters of each tree namely, DBH, height, and species and this used to inject into allometric equation to convert these measurements to AGB. Allometric equations are equations relating the sizes and volumes of an organism, and in the case of trees are normally produced not from first principles, but from destructively harvesting trees.

DBH is the most useful parameter to include in an allometric equation for estimating AGB, as DBH correlates well with AGB in every environment, and is comparatively easy, quick and accurate to measure. Height is of secondary importance, helping to refine the accuracy of the estimate derived from DBH. Height is especially important when using allometric equations not specifically derived at the same location and with the same species being studied, as much of the error between different allometric equations involving DBH alone can be explained by changes in DBH-height relationships across different areas (Chave et al. 2005). Finally, knowing the species is important because it allows the inclusion if a species-specific wood gravity (also known as wood density) measurement. Wood specific gravity is defined as the over-dry mass of a sample wood (in grams) divided by its green volume (cm^3). Height and diameter essentially just allow the estimation of a tree's volume (V): wood specific gravity allows this to be accurately converted to biomass, again increasing accuracy.

The importance of the choice of allometric equation cannot be overstated. Different allometric equations can produce very different AGB estimates using the same measurements, especially for large trees (> 50 cm DBH) (Keller et al. 2001; Williams et al. 2008). For the vast majority of tropical tree species, insufficient destructive harvesting measurements have been made to allow for species-specific relationships to derived (Cole & Ewel 2006). Even if such relationships did exits, some would argue that using them could introduce errors in a different site, or if the equation was derived from relatively few trees (Henry et al. 2010).

Given the impossibility of using species-specific equations in most circumstances in the tropics, it is normal to use pan-tropical equations, which collate thousands of individual measurements in order to derive generally-applicable equations, with known uncertainties. The equation most widely used in the tropics was derived by Jerome Chave (Chave *et al.* 2005), who collated 2,410 individual tree measurements to derive equations (involving either DBH and wood density, or DBH, height and wood density) for dry, moist, wet and mangrove tropical forests. These equations are widely trusted and used, and are thought to be relatively accurate (~ ±10 %), especially in the form including site-specific height measurements. The allometric equation model by Chave et al. 2005 has been improved by including twice the number of trees from 58 sites around the tropics and has been updated to the most recent years (Chave et al. 2014). This dataset includes the tree data from Malaysia and the Pasoh Forest Reserve (PFR) is one of the test site used to develop the equations developed by Jerome Chave (Chave et al. 2005, 2014).

2.3 Airborne LiDAR

2.3.1 An overview of airborne LiDAR

The LiDAR sensor emits pulses of light to determine the range (or distance) to a target object (Lim et al., 2003). The distance to the target is measured by the elapsed time between the emission of the pulse and the detection of the reflected (or backscattered) signal (Gatziolis and Anderson, 2008):

$$\text{Range (m)} = (\text{Speed of Light} \times \text{Time of Flight}) \div 2 \quad . \qquad (2.1)$$

The coordinates (x, y, z) of the target object is defined with the known position and orientation of the sensor, and the range measurement.

An airborne LiDAR system comprises a set of instrument: a laser range unit, an inertial navigational measurement unit (IMU), a high-precision global positioning system (GPS) receiver and a computer interface (Figure 2.1). The GPS and IMU continuously record the three-dimensional (3D) position and attitude (i.e., roll, pitch and yaw) of the platform, and the computer interface manages communication among devices and data storage (Gatziolis and Andersen, 2008). Airborne LiDAR systems can be categorized as either discrete-return (DR) or full waveform (Figure 2.2), and differ with respect to how backscattered laser energy is quantified and recorded by the system's receiver (Gatziolis & Andersen, 2008). Full waveform records the time varying intensity of the returned energy from each laser pulse, providing a record of the height distribution of the surfaces illuminated by the laser pulse (Lefsky et al. 2002). DR systems quantify

reflected energy (at amplitude intervals) and recorded return targets referenced in time and space (White et al., 2013). In a forested area, a laser pulse may be reflected from several surfaces, e.g., the canopy (i.e., branches and leaves) and often the ground. The resulting dataset of discrete-return LiDAR is a cloud of 3D points, with the upper points representing the canopy and the lower points representing the ground (Lim et al., 2003). As a mainstream LiDAR system for various environmental applications, airborne DR LiDAR has been used in the large number of studies for biomass and forest attributes measurement (Chen et al. 2013).

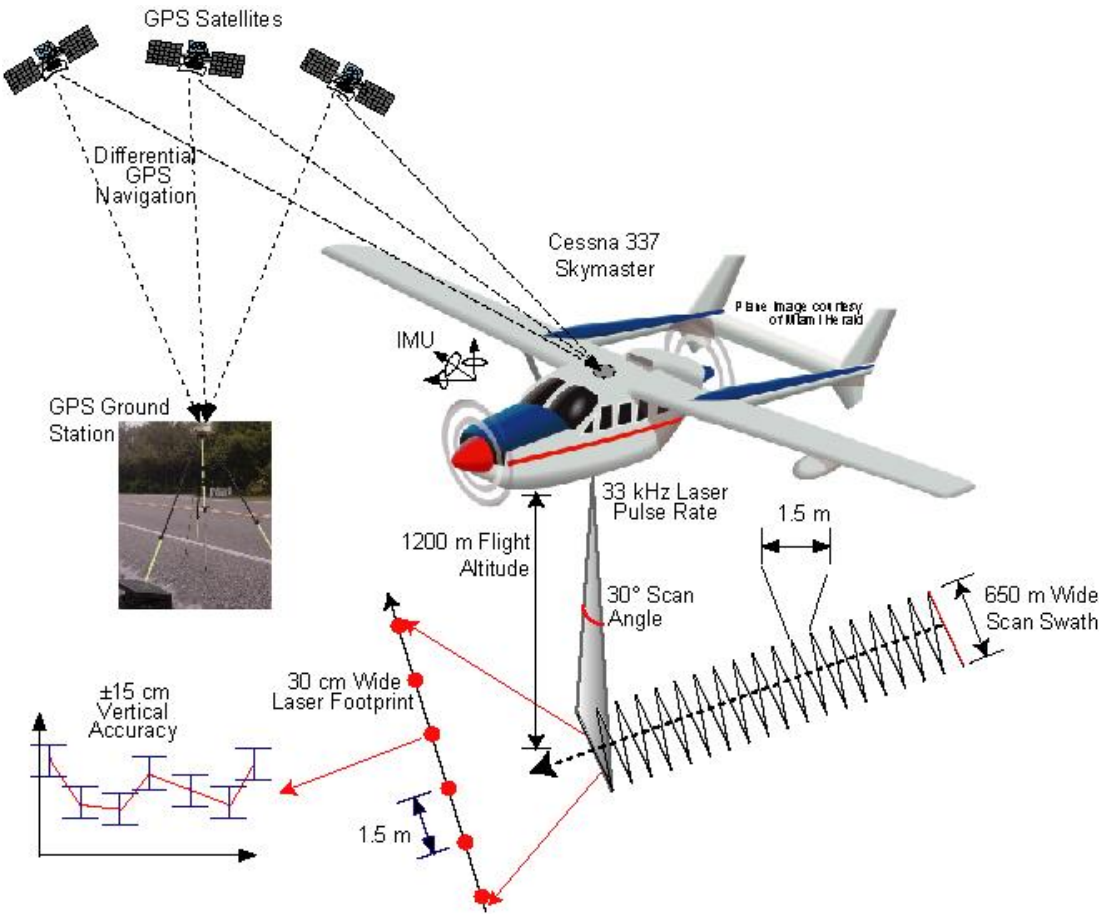


Figure 2.1. Schematic diagram of an airborne LiDAR system acquisition. Adapted from LiDAR technology, Retrieved June 30, 2017, from

http://lidar.ihrc.fiu.edu/aboutlidar.html. Copyright 2017 by International Hurricane Research Center.

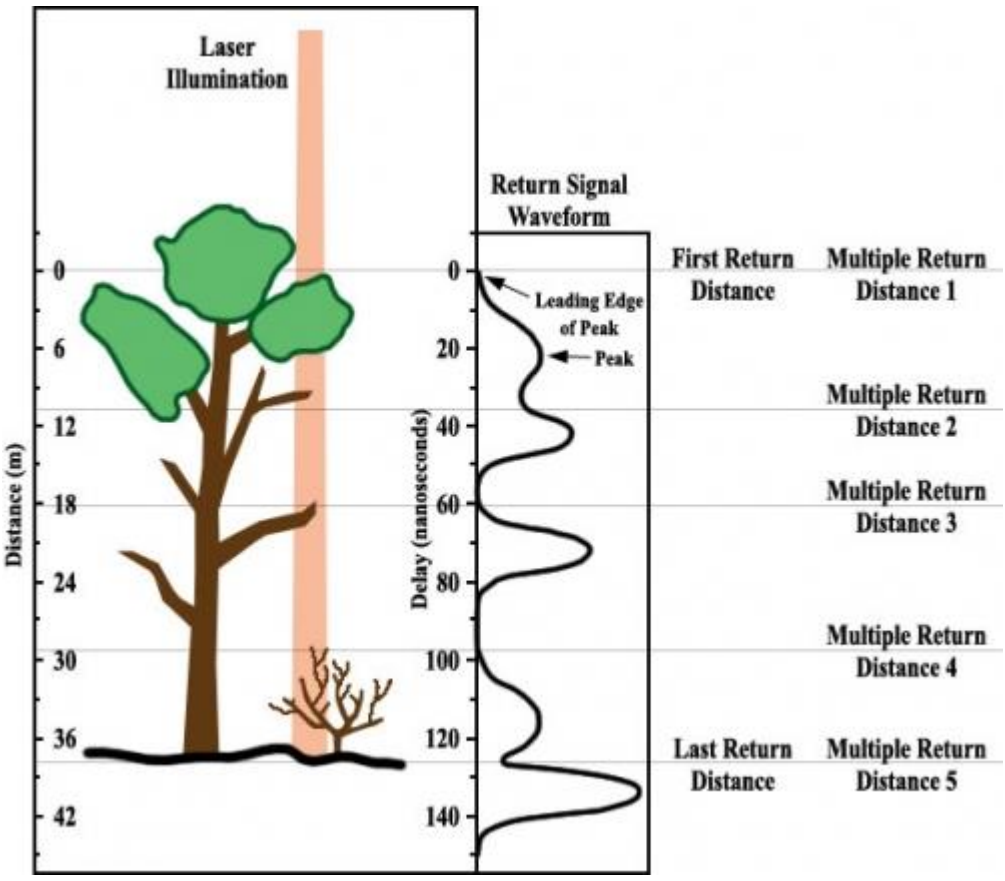


Figure 2.2. Graphical illustration for conceptual differences between waveform and discrete-return LiDAR. Adapted from "LiDAR for Biomass Estimation", by Vazirabad Y.F & Karslioglu M.O. *Biomass-Detection, Production and Usage.* 2011. Copyright InTech.

2.4 Airborne DR LiDAR for forest measurement

The use of LiDAR data in forestry has seen steady growth over the past several decades. The usefulness of LiDAR can be demonstrated specifically for forest

measurements (Zhao et al. 2011; Packalén et al. 2013; Dalponte et al. 2014) and estimation of forest biophysical parameters (Popescu et al. 2011; Hudak et al. 2012; Naesset et al. 2013) such as tree height, crown dimensions and tree location at the stand level, plot level and tree level. These direct LiDAR measurements can then be used in conjunction with known allometric relationships or statistical analysis procedures to estimate parameters such as DBH, AGB and V. LiDAR research for forestry applications has largely focused on the development of methodologies to employ LiDAR data as a substitute for various ground measurements.

LiDAR data can be collected over larger areas with a reduced amount of effort compared to traditional field measurements. However, the high level of complexity present within many forests (e.g., large number of species and variable canopy densities) can complicate the retrieval of such measurements. Since LiDAR systems collect data looking down on the forest, forest measurements other than tree height or crown dimensions (e.g., DBH, AGB) are typically indirectly estimated. Popescu (2007), used regression analysis to estimate the DBH of individual trees, using the LiDAR-derived height and crown width measurements as independent variables in a regression analysis. In traditional forestry, biomass estimation requires destructive sampling, or the use of species-specific (Chave et al. 2005; 2014) allometric equations. Allometric equations can also be applied to LiDAR data, if the required information is available. Popescu (2007) outlined a method for obtaining individual tree AGB estimates using allometric equations and estimates of individual tree DBH from LiDAR data. Examples of other studies that have also predicted AGB using LiDAR data include Lefsky et al. 1999; Patenaude et al. 2004 and Zhao et al. 2009. Forest

Research Institute of Malaysia (FRIM) has started to follow the initiatives runs by The United States Forest Service (USFS) Forest Inventory and Analysis (FIA) program to provide an inventory measurements to assess the status of the nation's forest. Forest resource managers and researchers commonly use these measurements to estimate forest biophysical parameters such as, V, AGB, or Carbon stocks (C).

With the wide introduction of LiDAR into remote sensing forestry studies, an increasing number of studies have undertaken individual tree detection (Kaartinen et al. 2012). LiDAR applications for biomass and forest measurements can be divided into two main categories: area based approaches and individual tree based approaches. In area based approaches, metrics derived from LiDAR point cloud are used to predict forest characteristics such as mean tree height, mean diameter, basal area, volume and biomass at the plot, stand level or for other areas of interest, typically using discriminant analysis, regression or non-parametric estimation techniques. However, this method is strongly based on accurate field data.

While, the basic principle of the individual tree based methods are to automatically identify individual tree crowns and extract individual tree information such as tree height, crown size and tree species using statistical techniques and allometric equations to derive individual tree attributes such as biomass, volume, basal area and diameter-at-breast-height. A specific method is required to delineate individual tree boundaries. Once correctly delineated, LiDAR statistical metrics such as those in area based approaches can be extracted by analysing the LiDAR point cloud within individual tree polygons. Compared to area based approaches, theoretically no fieldwork is

needed for model development once the individual tree crowns can be delineated with high accuracy or no errors and statistical model or allometric model exist to predict forest attributes based on LiDAR metrics. If there are no such model, the amount of fieldwork required is still much smaller than that for area based approaches, because field sample data are needed only for a sample of trees instead of a sample of plots or stands. Individual tree based approaches are more prominent than area based approaches in terms of making the LiDAR inventory more closely resemble traditional field-based inventories but at much larger scales, improving inventory precision that include listed attributes of each tree and making the classification of tree species much easier (Vastaranta et al. 2009). These studies have shown increased complexity of analyses, increased accuracy of results and a focus on the use of LiDAR data alone.

2.5 Approaches utilising LiDAR data for individual tree crown detection and delineation

Individual tree crown (ITC) delineation method using LiDAR data typically utilize a canopy height model (CHM) or digital surface model (DSM) and assume the tree top is the point with local maximum height value. Crown boundaries are assumed to be outlines with minimum height value (Borgefors et al. 1999; Zhen et al. 2014). ITC delineation approaches have been greatly enriched by a number of 3D methods using LiDAR data, which Koch et al. 2014 grouped into four categories: (1) raster-based methods; (2) point cloud based methods; (3) methods combining raster, point, and a priori information; and (4) tree shape reconstruction methods. The characteristics of four methods using LiDAR data in ITC delineation can be summarized in Table 2.1. The precise details of methods used will be discussed in more detail in later chapters,

but this section provides a brief introduction to the choice of method for ITC delineation. A focus given here for a method combining raster, points and *a priori* information.

Table 2.1. The characteristics of the four methods using LiDAR data for ITC Delineation. Adapted from "Trends in Automatic Individual Tree Crown Detection and Delineation – Evolution of LiDAR Data", by Zhen Z., L.J. Quackenbush & Lianjun Z. *Remote Sens.* 2016, 9, 333. Copyright 2016 by MDPI.

Method Group	**Algorithms/Methods**	**Advantages**	**Disadvantages**
Raster-based	Tree top detection: local maximum, image binarization, template matching Crown delineation: valley-following, region-growing, watershed segmentation GEOBIA-based method	• Well developed • Easy to be used and improved • Easy to use multiple data sources if using GEOBIA-based method • Easy to use multiple scales if using GEOBIA-based method.	• Missing information or causing potential errors from extraction, interpolation and smoothing procedures
Point cloud-based	Clustering technique (K-means, Mean shift, Hierarchical) Voxel-based single tree segmentation	• Easy to use 3D information • Easy to reflect canopy structure • Could detect understory trees or small trees	• Harder to implement than raster-based method • Greatly depend on point density of LiDAR data
Methods combining	Classic ITC delineation algorithms + prior		

raster, point and *a priori* information	information (crown size/stand density) Imagery + point cloud	• Benefit to use historical data • Easy to integrate remotely sensed data and GIS data • Could use both spectral and height information	• Depend on prior information • May have difficulties in registration of different data sources
Tree Shape reconstruction	Convex hull Alpha shape Superquadrics Hough transform	• Further delineation of crown • Provide 3D individual tree profiles • Provide information for estimating foliage and photosynthetic activity • Allow further application of modelling tree growth	• Difficult to implement • Sometimes depend on successful segmentation of single trees • Difficult to collect field data for validation

*GEOBIA = Geographical object-based image analysis

2.5.1 ITC delineation through combining raster, point and *a priori* Information

Raster-based methods have become the dominant method for LiDAR inputs and have a longer development history than other approaches. Local maximum, template matching, region-growing and watershed segmentation is an example of all traditional tree detection and delineation algorithms than can be applied on a raster CHM, which is extracted from the laser point cloud, interpolated and smoothed (Koch et al. 2014).

In most of the traditional approaches, tree detection is generally based on finding local maxima within the image and crown delineation requires outlining minimum valleys regardless of whether the input is a passive image or a CHM or DSM raster (Wulder et al. 2000; Hyyppäet al. 2001; Chen et al. 2006). As development through years, raster-based methods have been incrementally advanced by addressing a specific issue (Gleason and Im 2012; Zhen et al. 2015) and incorporating geographic object-based image analysis (GEOBIA) approach (Suarez et al. 2005). Since raster-based method using active data that require extraction, interpolation and smoothing, these all procedures have the potential to cause errors. Thus, point-cloud methods came as an alternative approach to avoid the potential errors.

Most of the point cloud-based methods directly apply LiDAR point data. Voxel-based single tree segmentation and clustering (e.g. K-means) techniques are one of the point-cloud based methods. A volumetric pixel or voxel is a representation of LiDAR data, also a basic 3-Dimensional (3D) element for exploring canopy structures, has been frequently used in ITC delineation related studies (Popescu & Zhao 2008 ; Wang et al. 2008). Voxel-based methods provide a convenient way to reflect canopy structure, but this method tend to greatly impacted by the density of the LiDAR points. Clustering method such as k-means, is a popular iterative partitioning approach, normally requires seed points and then partitions LiDAR points into a group of clusters using a distance criterion (Gupta et al. 2010). Li et al. 2012, designed 3D point-based method that does not relies on data training to perform segmentation. Instead, they used the highest points within a threshold distance as seed points. Their algorithm exploits the spacing between the tops of trees to identify group points into a single tree based on simple

rules of proximity and likely tree shape. This method worked well for coniferous types of forests, but to adapt this method at different forest site a few settings are recommended such as applying adaptive threshold based on prior knowledge of study site.

With recent advances in information content and richness of data, researchers have explored the benefit of integrating raster, point and *a priori* information in ITC delineation studies. The integration of methods usually conducted using multiple approaches. Adaptation of ITC delineation algorithms by taking advantage of prior information for segmentation (like stand density) and a methods combining image and point cloud analyses (Koch et al. 2014) are one of those integrations approaches used by researchers. The most useful information that can be incorporated into ITC delineation studies is the crown size and stand density (Koch et al. 2014; Zhen et al. 2015). Heinzel et al. (2011), introduced a prior knowledge based watershed segmentation that first classified crown sizes using iterative granulometry and then delineated tree crowns using watershed segmentation. Chen et al (2006) and Zhen et al. (2014) employed local maxima methods with variable window size to detect treetops based on crown size estimates. Ene et al. (2012) generated CHMs with variable resolutions and selected the most feasible CHM using area-based stem number estimates to guide filter size. Hauglin et al (2014) designed an adaptive approach that initially delineated crown from a CHM using a marker-based watershed algorithm and then guided final delineation using a priori area-based stem number predictions. Brandtberg et al. 2003; Falkowski et al. 2006 and Jing et al. 2012, implemented adaptive parameterisation approach incorporates multi-scale analysis.

The detection of tree crowns is based on smoothing the rasterized data at multiple scales through a range of image processing techniques (e.g. Gaussian filters) with specific scales depending on the size of tree crowns. Jing et al. 2012, analysed a set of similar things within a LiDAR CHM through a series of morphological opening operations (Soille 1999) to define the range and dominant sizes of crowns to guide the design of the Gaussian filters before applying multi-scale filtering. This approach reported performed well when compared with manual interpretation.

Methods that combine image and point cloud analyses are frequently used from a data-integration perspective. Reitberger et al. (2009), for instance, segmented individual tree crowns using conventional watershed algorithm based on CHM and followed this with a 3D segmentation of single trees using normalized cut segmentation based on point cloud data for segmenting small tress that fell below the CHM. The integration methods of watershed-based ITC delineation of a CHM and refinement of the results using LiDAR point cloud performed well not only for dominant and codominant trees, but also for intermediate and suppressed trees. The trough algorithm developed by Duncanson et al. 2014 applied to the LiDAR point cloud worked better to detect understory crowns in the conifer site than in the closed canopy deciduous site (Duncanson et al. 2014).

2.5.2 Selection of suitable variables from LiDAR data for AGB estimation forest parameters modelling

A wide range of LiDAR metrics have been used for AGB estimation in the literature (Chen 2013; Maltamo et al. 2014). Extraction of LiDAR metrics depends on the laser

return signal (discrete-return vs. waveform), scanning pattern (scanning or profiling) and footprint size (small vs large). Since discrete-return LiDAR used as a main tool and used throughout this thesis, the author only discuss the metrics from these systems. In airborne LiDAR data, metrics can be extracted on the basis of either individual trees or areas (Chen 2013). The individual tree based approach requires identifying tree features such as treetop (Chen et al. 2006), crown radius (Popescu et al. 2003) or crown boundary (Zhen et al. 2014). Once an individual tree crown identified, information from individual trees such as tree height and crown size which can be related to biomass and other forest structures variables can be extracted through allometric equations (Chen et al.2006; Chen et al. 2007). Theoretically, no fieldwork is needed for model development if tree crowns can be delineated with no errors and strong allometric equations exist to estimate AGB based on LiDAR metrics such as tree height or crown size. Even if there are no such allometric equations, the amount of fieldwork required is still much smaller than that for area based approaches needed because field data needed only for a sample of trees instead of a sample of plots or stands like in area-based approach needed (Chen 2013).

The mostly widely used LiDAR metrics for AGB prediction are various height metrics (Lim & Treitz 2004; Patenaude et al. 2004; He et al. 2013; Wan-Mohd-Jaafar et al. 2017). LiDAR metrics can be calculated based on first, last or all returns (Chen 2013). The best height metric reported in the literature differs a lot in previous studies, including the 80^{th} (Patenaude et al. 2004; Wan-Mohd-Jaafar et al. 2017), 75^{th} (St-Onge et al. 2008), 50^{th} (Thomas et al. 2006), 30^{th} (Stephens et al. 2012), and 25^{th} (Lim & Treitz 2004) percentile heights. The differences in vegetation structure, model

development and data processing procedure may result such variability but the dominant cause is unclear. It is important to test which LiDAR variables that interact very well with AGB according to different test site in order to get an accurate estimation. Beside height metrics, other metrics such as the one related to canopy cover, also known as crown cover (Hall et al. 2005), canopy density (Naesset & Gobakken et al. 2008) and canopy volume (Kim et al. 2009) have been proved to be useful for predicting AGB as well. Canopy cover and density can complement height metrics for a better characterization of 3D canopy structure because they are related to horizontal and vertical vegetation structures respectively. These metrics has been successfully used for predicting AGB (Hall et al. 2005), stem volume (V) and basal area (Chen et al. 2007).

2.6 Parametric and non-parametric approaches for modelling AGB and forest parameters using Airborne LiDAR data

Airborne LiDAR has demonstrated an ability to improve the accuracy of forest inventory parameters such as height (Ht), stem volume (V) and basal area (BA) estimates. Scaling the forest inventory information acquired from field measurement or remotely sensed data requires statistical modelling to obtain wall-to-wall information from discretely sampled observations. Various methods have been implemented by using LiDAR data for modelling AGB and other useful forest parameters and a brief introduction to the choice of method for modelling forest attributes using parametric and non-parametric approaches will be presented here.

Most of inventory predictions from LiDAR point cloud statistics have involved parametric regression. Parametric methods can be used to estimate selected variables of interest (Wang et al. 2005; Salas et al. 2010). Linear and non-linear models have been used for this purpose in previous studies (Nelson et al. 2004; Wang et al. 2005). Ordinary least square (OLS) regression is one of the type of linear regression modelling that has been used widely with LiDAR data and biomass estimation by most remote sensing analyst (Næsset & Gobakken 2008). Fassnacht et al. (2014) shows that the most common prediction methods in LiDAR-based forest inventories are ordinary least square (OLS) regression with stepwise variable selection. This method has been most frequently used for building models between field measurements and LiDAR metrics (Garcia-Gutierrez et al. 2014). The main advantage of using OLS methodology is because of the simplicity and clarity of the resulting models (Garcia-Gutierrez et al. 2015) especially when the relationship between response variables and LiDAR metrics is almost linear. This method minimizes the sum of squared vertical distance between the observed responses and the responses predicted by linear approximation. OLS shows competitiveness in estimation of forest biomass and carbon stocks using LiDAR data compared with non-parametric methods (Li et al. 2014).

Non-parametric methods are an alternative to the OLS regression. K-nearest neighbour (*k*-NN) for example, is one of the non-parametric approaches have been used to performed multivariate analyses of forested areas by associating variables of interest (i.e. ground data) to LiDAR data (Hudak et al. 2008; Goerndt et al. 2010). Non-parametric algorithms have been recently explored for forest attributes estimation due to their flexibility and ability to describe nonlinear dependences compared to

parametric algorithms (McRoberts et al. 2007). One of the greatest advantages is that they are free assumption of any given probability distribution (Sironen et al. 2010). The methods that belong to the nearest neighbour categories: most similar neighbour (MSN) imputation (Muinonen et al. 2001; Kankare et al. 2013) and gradient nearest neighbour (GNN) (Hudak et al. 2008; Hudak et al. 2014; Pierce et al. 2009; Temesgen & Ver Hoef 2015) imputation. Random-Forest (RF) is one of the non-parametric approach that is widely used in LiDAR forest inventory by analyst (Hudak et al. 2008; Falkowski et al. 2010; Hudak et al. 2012; Temesgen & Ver Hoef 2015). MSN and GNN have been widely used in the United States region (Pacific Northwest) (Shin et al. 2016).

None of the parametric and non-parametric techniques exist in the study are consistently superior than another. Additional properties of the estimation methods such as assumption, ease of implementation and data requirements should be considered in order to assess the performance. Parametric methods for example, can give a good prediction for estimating forest attributes from LiDAR with a relatively small sample size, while non-parametric techniques are heavily reliant on sample selection and adequate sample size. Results in non-parametric method are generally evaluated in terms of the predictions and the predictions are less sensitive to changes in data calibration. While in parametric methods, all predictions are affected. Parametric methods is limited in terms of implementation as the approach requires stratification by forest type, while no stratification framework by forest type required in non-parametric approaches. Non-parametric methods can produce relatively more robust model as it uses cross-validation during model development, this is considered

as one of the important advantages that differ non-parametric then parametric. The expertise and time required to implement and develop parametric and non-parametric methods may vary. Both benefit from increased the statistical knowledge but parametric approaches generally benefit more in terms of model selection and fitting routines.

CHAPTER 3

The Study Area: Pasoh Forest Reserve and FRIM Forest Reserve

3.1 Introduction

The goal of this chapter is to give an overview of the study areas by describing in detail on the geographical location, climate, topography, land use management and history of the study areas. This chapter also describe in detail on how the field data were collected and the technical processing of field data. The analysis and the interpretation of the fieldwork data concludes the section. The acquisition and the specification of airborne LiDAR data used in this thesis described at the end of this chapter.

3.2 Forest Site

3.2.1 Pasoh Forest Reserve (PFR)

The PFR is located on peninsular Malaysia, just 140 km southeast of Malaysia's highly populated capital city, Kuala Lumpur (Figure 3.1). The PFR study site (2.98 N 102.31 E) is located about 8 km from the town of Simpang Pertang, in the state of Negeri Sembilan. PFR has an area of approximately 140 km^2, mainly covered with lowland dipterocarp forest and with hill dipterocarp forests in its north-eastern boundary. The core area which is approximately 600 ha within the reserve is still covered with old

growth forest; most of the surrounding area has been logged in the past, providing many examples of regenerating lowland forest. PFR is one of the most complex dense species-rich communities in the world with 340,000 trees consisting of 818 species.

This reserve forest is under the custodian of the State Forestry Department of Negeri Sembilan, Malaysia. In 1977, it was designated as a forest reserve to function for a variety of purposes, including production forestry (sustainable timber harvesting) and the protection of flora and fauna (conservation and research). FRIM was given the mandate to manage the forest area and facilities for research purposes. Previously it was under the management of Malaysia's first university, University of Malaya, which was conducting pioneer research work with researchers from United Kingdom under the International Biology Programme with UNESCO and the Man and Biosphere (MAB) programme in 1971. Since FRIM took over its management in 1977, the forest research area has gained an international reputation as an ideal field laboratory for tropical forest ecological studies (Fletcher *et al.* 2012).

Within this reserve is an area of 18.4 km^2 set aside for research purposes. The core area of 4 km^2 consists of pristine forests, while the remaining areas are regenerating forests harvested for timber in the 1950s. Established in 1985 together with the Center for Tropical Forest Science (CTFS) under the Arnold Arboretum and Harvard University, for research purpose, the Malaysia forest authority has maintained a 50 hectares permanent tree demographic plot and censussed every five years. All stems (trees and plants) 1 cm at diameter breast height and above are mapped, tagged,

identified, measured and counted using the CTFS methodological standards. FRIM also maintained the five 2-ha permanent ecological plot established in 1971. All trees ≥ 5 cm diameter at breast height have been mapped, tagged and identified and the 2-ha arboretum with 874 tagged and mapped trees belonging to 279 species from 44 families.

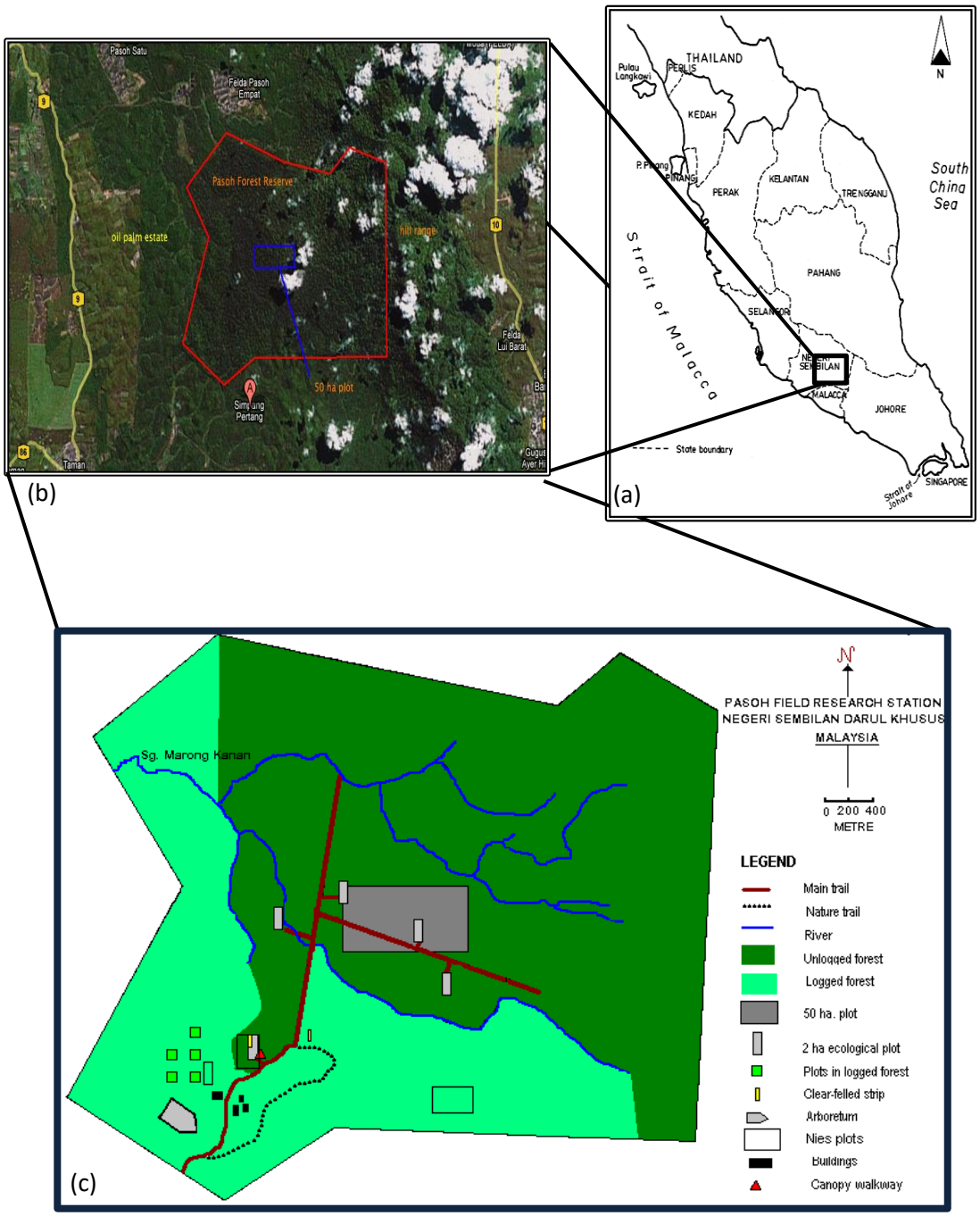


Figure 3.1. Study area; (a) location map, (b) Approximate bounds of the forest reserve from Google Earth Image, (c) PFR site location, administration and scientific infrastructure. Adapted from Forest Research Institute Malaysia, Retrieved March 1, 2014, from https://www.frim.gov.my/en/about-us/field-research-stations/pasoh-frim-research-station-negeri-sembilan/frimctfsaa-harvard-pasoh-50-ha-plot-recencus. Copyright 2014 by Forest Research Institute Malaysia.

3.2.2 FRIM Forest Reserve

FRIM is located in Kepong, near Kuala Lumpur, and extends from about 54 m above mean sea level (ASL) and gradually increase in the hilly terrain to reach the highest peak at about 300 m height ASL. FRIM located 140 kilometres away from the main test site of PFR (Figure 3.2). FRIM Forest Reserve receives rain throughout the year at the average rainfall of about 2,700 mm and the average temperature is about 32^0C. FRIM Forest Reserve can be considered as among the last remaining green belt in Kuala Lumpur.

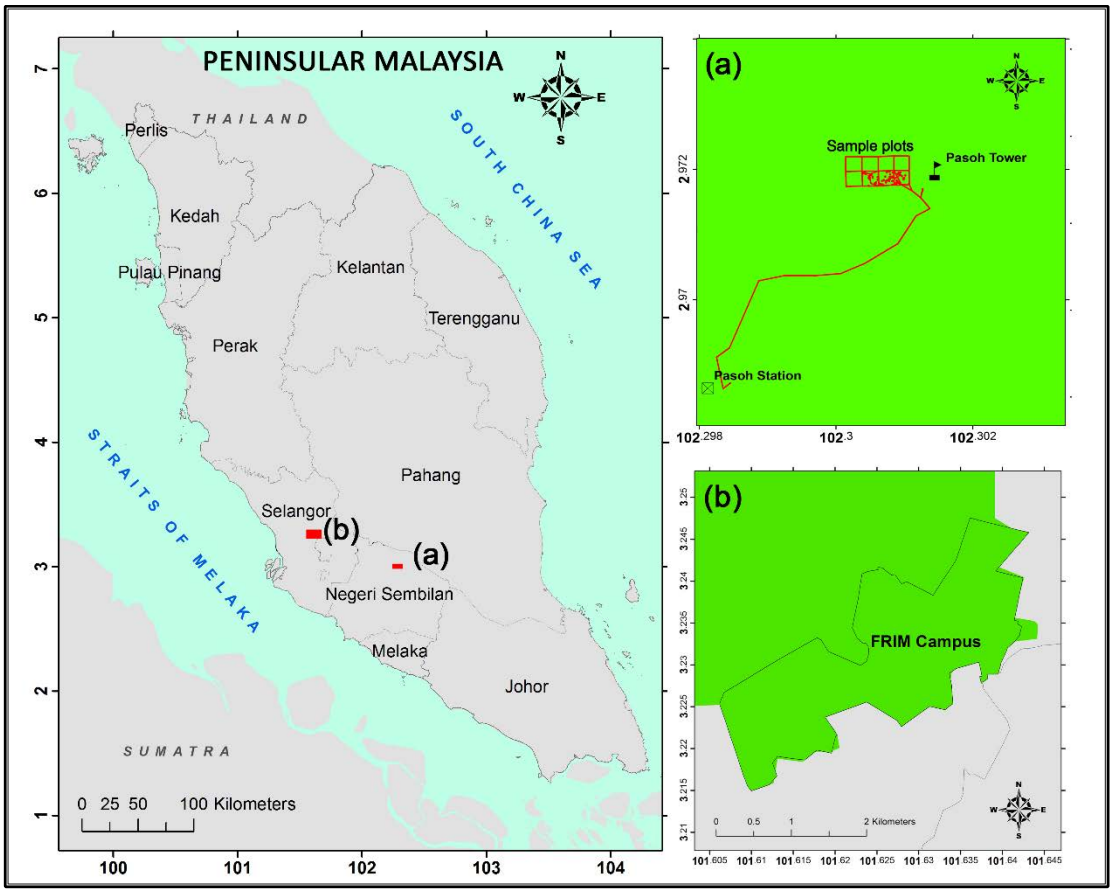


Figure 3.2. Study area; (a) PFR, 6-ha IBP plot location and (b) 600-ha FRIM located inside FRIM campus.

The early site comprised an area that was practically stripped of its original forest cover except for a few remnant trees at more inaccessible localities. Major plantation research began in FRIM campus in the 1920's. More than 100 species both natives and exotics were planted in 1920's and 30's (Barlow 2010). The planting was successful, however no scaling up of the efforts towards industrial tree plantation levels were embarked. As a result, these areas serve as the demonstration sites rather than production areas. Over the years since 1970's several plantation trials of exotic species such as pines, *Eucalyptus spp, Araucaria* spp *and Acacia* spp were established to support wood based industry. The establishment of all these plantations trials together with existing native forest within the FRIM campus and its close proximity to the permanent reserve forest makes it a unique setting for natural succession of flora and fauna species into the area. Some of the planted forest stands have grown to a fully stocked stand approaching perhaps the condition of the pristine climax forest. Fortunately, one of the long-term ecological research plot of old growth hill dipterocarp forest was within the adjoining Bukit Lagong Forest Reserve. Comparative assessment of the planted trees with the pristine forest will indicate the ecological integrity of the forest ecosystem using some simplified metrics indicators that will indicate its status. FRIM's initial investigation on the several matured stand types indicated that the stand had the form of an inverse J-shape size distribution common to many uneven-aged and multi species stand of tropical mixed forest (Kassim 2015). In general, the stand and stock tables indicates that the stands had grown to fully stocked stands comparable to the highly productive stands of hill dipterocarp forest. Some species have grown to a maximum size for native trees, with heights of 60 m and dbh of 164 cm equivalent to mature trees in the natural forest.

3.3 Vegetation

3.3.1 PFR Vegetation

In the main PFR 50-ha plot (Table 3.1) the *Dipterocarpaceae* family dominates, accounting for 27% of the total basal area (Manokaran & LaFrankie 1990). The *Euphorbiaceae* was the richest family in the plot with 85 species and had the highest number of trees in the plot with 13.4% of total tree numbers. *Shorea* was the most important genus in the 5-ha research plot in terms of tree number where it contributes 6.2% of all trees (20,960 trees) and it was the fifth most diverse genus in the plot with 14 species (1.7% of all species). The family *Dipterocarpaceae* once again dominated the plot with 10 common genuses for timber groups. *Red Meranti* was the biggest genus under *Shorea* with a total of 13, 401 trees (43.36% of total basal area) followed by *Balau* under the same genus with 6842 trees contributed 22% of the total basal area (Davies *et al.*, 2003).

The fieldwork in this thesis was conducted within the 6-ha International Biological Programme (IBP) plot, known as "the ecological plot", in which all trees (≥ 5cm in DBH) have been measured and mapped since the early 1970s; the species composition, stand structure, and other vegetation properties of this plot have also been monitored (Manokaran 1998). The species dominated in this plot was a bit similar to the 50-ha plot (Table 3.2).

Table 3.1. Important families for 50-ha plot of PFR

Family	Basal Area (m^2)	Family	Total number of Trees	Family	Species number
Dipterocapaceae	453.21	Euphorbiaceae	45436	Euphorbiaceae	85
Fabaceae	141.47	Dipterocarpaceae	31178	Lauraceae	49
Euphorbiaceae	120.46	Annoaceae	24752	Myrtaceae	48
Burseraceae	100.91	Rubiaceae	20506	Rubiaceae	47
Myrtaceae	56.96	Burseraceae	17701	Annoaceae	42

Source: Davies et al., (2003)

Table 3.2. Original composition of dominant tree species (DBH ≥ 10 cm) in the ecological plot (20m x 100m) taken in 1994 census

Family	**Species**	**N**	**Max DBH (cm)**
Leguminose	Koompassia malaccensis	1	100.9
Dipterocarpaceae	Dipterocarpus cornutus, Shorea Macroptera, Shorea dasyphylla, Shorea paucliflora, Shorea multiflora	9	77.6
Sapotaceae	Ganua	8	32.8
Guttiferae	Mesua	1	65.6
Sapindaceae	Nephelium ramboutan	2	53.2
Anacardiaceae	Pentaspadon motleyi	2	57.9
Leguminosae	Sindora sp., Dialium procerum	4	59.8
Euphorbiaceae	Neoscortechinia kingii, Pimelodendron griffithianum	3	52.5
Myrtaceae	Syzgium hoseanum	2	42.0
Myristicaceae	Knema sp, Gymnacranthera farquhariana	2	45.7
Oxalidaceae	Sarcotheca griffithii	1	44.8
Meliaceae	Lansium domesticum Correa	4	34.1
Flacourtiaceae	Ryparosa wallichii	1	38.1

Source: FRIM

3.3.2 PFR Forest Structure and species composition

The most prominent feature in this reserve forest is the emergent trees whereby the crown of these trees grow out of the main forest canopy (Figure 3.3). The emergent trees in this reserve are mainly dipterocarp trees (*Shorea spp. – Meranti, Dipterocarpus spp.-keruing* and *Neobalanocarpus heimii-Chengal*) and the legume family, Leguminosae (*Koompasia malaccensis – Kempas tree and Intsia palembanica – Merbau*) among others. These trees can grow to impressive heights of over 60 m tall and girth of over 2 m. The biggest tree recorded within the PFR demography permanent ecology plot is a Chengal tree measuring 231.2 cm DBH. Beneath this emergent layer is the main canopy and lower-storey trees consisting of more general and commonly distributed tree families such as *Myrtaceae (Kelat), Guttiferae (Mangosteen), Sapindaceae (Rambutan) and Anacardiaceae (Mango).* Many of these trees produce edible fruits. The ground vegetation consists of smaller plants like gingers, palms, ferns and herbaceous plants.

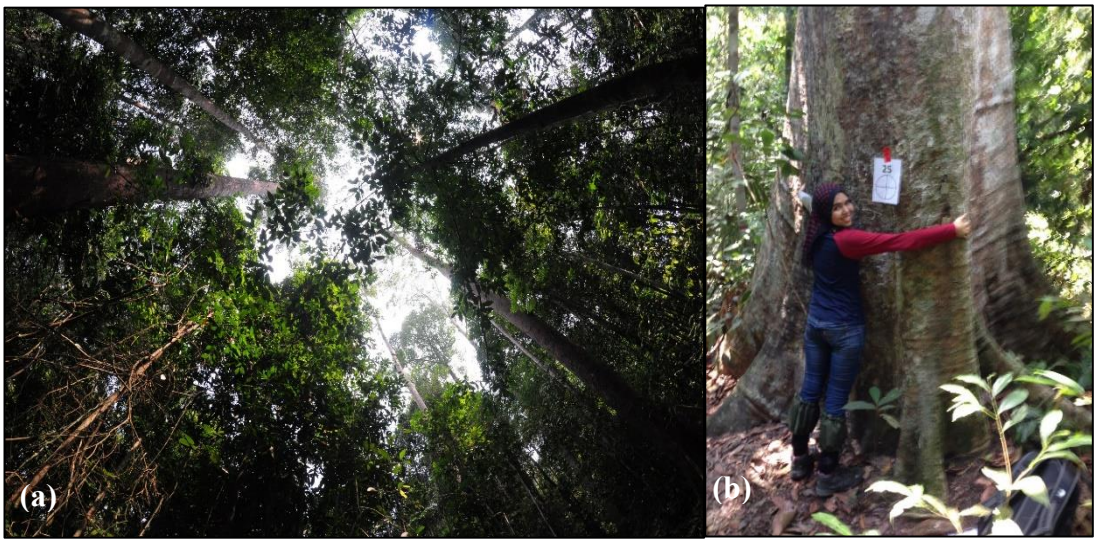


Figure 3.3. (a) tree structure in PFR from ground view taken during fieldwork campaign and (b) 100 years old tree tree spotted in plot A3, *Koompassia malaccensis*, with DBH 109.9 cm and 32.8 m tall.

3.3.3 FRIM Vegetation and Species Composition

Early years, forestry research focused on evaluating the performance of several species to restore degraded lands with high quality of local and exotic species for timber production. Among the species that have been planted are *Anisoptera scaphulata*, *Neobalanocarpus hemii, Dryobalonops aromatica, Dryobalonops oblongifolia, Dipterocarpus grandiflorus, D. costulatus, D. baudii, D. dyeri, D. cornutus, D. crinitus, D. kerrii, D. hasseilii, D. semivestitus, Hopea sangal, H. subulata, Shorea ovalis, S. curtisii, S. leprosula, S. hemsleyana, S. acuminta, S. macrophylla, S. parvifolia, S. pauciflora, S. hypochra, S. lepidota, S. resinosa, S. platyclados, S. glauca, S. collina, S. laevis, S. maxwelliana, S. ovalis, S. materialis, S. sumatrana, S. foxworthyi, S. rugosa* and *Vatica* spp. There are also several dipterocarp species that came from Borneo such as *Dryobalanops lanceolata* and *Shorea mecistopteryx*.

While for non-dipterocarp, the major species including *Acacia mangium*, *Paraserianthes falcataria, Araucaria cunninghamii, Araucaria hunsteinii, Aquilaria malaccensis, Azadirachta excelsa, Fagraea crenulata, Dyera costulata, Endospermum diadenum, Eucalyptus torrelliana*, *Fagraea fragrans*, *Khaya ivorensis, Kompassia malaccensis, Elateriospermum tapos*, *Agathis alba, Ochanostachys amentacea, Mesua ferrea, Calophyllum inophyllum, Strombosia javanica, Scorodocarpus borneensis, Palaquium gutta, Artocarpus nitidus, Styrax benzoin, Swietenia macrophylla, Eusideroxylon zwageri, Pentaspadon motleyi and Pinus insularis.* Other than tree species for plantation, FRIM also has a unique biodiversity collection of other flora including 160 species of mushrooms which are *Amanita* angustilamellata, *Boletus langipes, Russula virescens, Cookeina sulcipes* and others.

For ferns there are *Selaginella willdenowii, Adiantum latiflolium, Angiopteris evecta, Pleocnemia irregularis, Drynaria quercifolia,* and *Pyrrosia lanceolata.*

Among the unique species is *Bayabusua clarkei* which classify as a climber. Also recorded are *Begonia phoeniogramma, Acranthera didymocarpus* and *Licuala pusilla* which is endemic to Selangor, Negeri Sembilan and Peninsular Malaysia respectively. *Orchidantha fimbriata* and *Thottea tricornis* is also endemic and live naturally at FRIM's ground. It was believed that FRIM is one of the earliest locations in Malaysia to plan bamboo, which started in 1927, including giant bamboo, *Dendrocalamus gigantea*. FRIM also has specimens of species that were classified as threatened plants, or near threatened, under the International Union for Conservation of Nature (IUCN) classification. For example, *Podocarpus rumphii* found in FRIM is very rare in Peninsular Malaysia and is only available in Kedah, Pahang and Selangor. In its original habitat, it grows on a hillside or ridge up to an altitude of 1750 m at sea level. Another species that is listed under near threatened is *Podocarpus teysmanii* which can be found at Coniferatum. Most of the species and plants have been planted more than 100 years ago, meaning the generation of these trees is considered as pristine (Commissioner of Heritage, 2015).

3.4 Why PFR and FRIM?

The selection of the study area was based on: (1) the variety of vegetation types and (2) the availability of Airborne LiDAR data over the study area. The variety of the vegetation species that exist in the 6-ha plot of PFR and FRIM is enough to represent any tropical rain forest available within this region and suitable to be selected as a

study area due to its completeness of secondary data. It is not easy to acquire airborne remote sensing data for tropical rainforest sites with a ready and complete set of secondary data. A large number of biological research projects have been carried out at PFR since 1970s, which also helps when performing the literature search (Kato *et al.* 1978; Okuda *et al.* 2003; Chave *et al.* 2005)

3.5 Climate and Rainfall

3.5.1 PFR Climate and Rainfall

The climate of the PFR area is the driest and hottest of the southern peninsular of Malaysia. Rainfall ranges from 1,728 to 3,112 mm (mean = 2,054 mm), which is relatively low for Malaysia. A fairly even distribution of rain throughout the year nevertheless permits the development of typical lowland rain forest. Being an isolated forest surrounded by oil palm estates, formation of large forest gaps by wind throw is a fairly common feature at PFR.

3.5.2 FRIM Climate and Rainfall

The mean annual maximum temperature in FRIM for the period of 1950 to 2003 was between 30.8° to 31.6 °C. The highest temperature of 31.6 was noted between 1950–1969. This could be attributed by the fact that the tree canopy cover was smaller with lower evapotranspiration rates. On the other hand, the mean annual minimum temperature was between 22.2–23°C. The mean ten-year annual increase was 0.25°C and 0.27°C for maximum and minimum mean ten-year annual temperature

respectively. On the other hand, the maximum and minimum mean ten-year annual temperature for Klang Valley were 0.4–0.5 °C. It is therefore noted that human-made forest within a dense developed area, like FRIM, helps in ameliorating climate change impact. The temperature increase was below the annual mean (22.2–23°C) increase for Peninsular Malaysia. The mean annual rainfall for FRIM was 2,661 mm for period 1974–2009 and higher than the mean annual rainfall for Peninsular Malaysia, 2,000–2,400 mm for the period 1951–2014 (Commissioner of Heritage 2015).

3.6 Topography and Soil

3.6.1 PFR Topography and Soil

The forest area is relatively flat with undulating hills to the northeast, and reaches an altitude of 600 m above sea level. The main soil type consists of alluvium (in the lowland) and shale (on the hill slopes), with smaller areas of laterite (red, iron rich) soil on the hill summit. The alluvium soil contains more water as it is in the wet, swampy areas of the reserve, while the shale is more variable in texture. This provides a variety of habitat platform for a diversity of trees to establish in the reserve. PFR lies on a level plain of raised Pleistocene alluvium from which low undulating hills of Triassic sediments and granite arise. It is bordered to the east by a sharp north-south granite ridge that reaches a peak at Bukit Palong, 600 m above sea level (Fletcher et al. 2012). The PFR Forest Dynamic plot differs by only 24m in height from high to low point. While the plot contains no permanent streams, a significant portion lies under standing water for more than 1 month, typically during November and December.

3.6.2 FRIM Topography and Soil

In the early years of establishment of permanent FRIM campus in Kepong, the lowland flat and undulating terrains were mostly occupied with tin tailings soil, while the hilly and steep terrain were developed over the granitic rock parent material. Along this hilly terrain and with high precipitation, surface soil particle movement from hill top to the low-lying area formed colluviums while along the rivers and streams were alluvial soils. At present, FRIM constitutes 544.3 ha from low lying areas to hilly areas up to 30°. Soils are mainly formed from granitic materials, comprising common soil series such as Rengam, Tai Tak, Beserah, Batang Merbau, and Baling. The common series, Tai Tak and Rengam were fairly deep and had coarse to medium sandy clay texture, where clay values fall between 30- 55%. Most of the hilly areas are derived from granites, however the foothills comprise colluvial materials which is the mixture of weathered materials and smaller pebbles derived from granite. Soils in FRIM can be classified as moderately fertile for forest tree growth and was able to support tree growth up to 75 cm in diameter at breast height and 30 meters in height. (Wan-A-Kadir & Vijayanathan 2015).

3.7 Fieldwork Data Collection

3.7.1 Description of Fieldwork Measurement

Fieldwork was only conducted at PFR. In addition, pre-existing field data was assimilated from the FRIM site. Field data at FRIM was collected from the FRIM inventory work in early April 2014 from FRIM Geospatial Department. All the parameters collected and measurement guideline is same and at the same standard for

both study site. Fieldwork measurement was carried out from 28th to 30th October 2014 in PFR. The objective of the measurement campaign was to collect positional data and physical measurement data for the trees in PFR in the ecological field centre. The physical parameter comprised of latitude and longitude of individual tree position, total tree height, height of the first living branch or known as bole height, diameter at breast height (DBH), crown diameter and species information for each tree. Because of the challenge to get a signal from GPS devices under dense canopy, three different devices were used to measure positions to ensure an accurate position or at least close enough. Garmin Handheld and Trimble Laser Ace Range finder was used to measure individual tree position and Total station was used to measure edge of plot setup. The major steps involved, which are; (i) field plot setup and physical tree measurement (ii) data processing and tree mapping; elaborated in detail in the following sections and (iii) biomass calculation using allometry equations will be elaborated in the next section.

3.7.2 Sampling plots and physical tree measurement

To minimise the uncertainties associated with biomass measurements, the layout of forest sampling must follow the standard measurement guidelines from the Malaysia National Forest Inventory (MNFI). Stratified random sampling of a square shape with 50m x 100m (0.5-hactare) was split into 8 experimental plots with dimension of 25m x 25m each to measure trees with DBH > 10 cm for plot number A1, A2, A3 and A4 and the criteria increased for plot B1 to B4 where in this plot only emergent trees with DBH>10cm were taken into measurement (Figure 3.4). It was designed in such a way to facilitate the mobility of inventory work at the field as advised by FRIM officer.

There are 142 individual tree measurements recorded across all plots (Table 3.3) during the fieldwork campaign, but only 105 individual trees used for final assessment.

Table 3.3. Fieldwork sampling plot

Plot Number	Number of Trees	Criteria
A1	24	Trees with DBH > 10cm
A2	23	Trees with DBH > 10cm
A3	24	Trees with DBH > 10cm
A4	8	Trees with DBH > 10cm
B1	4	Emergent Trees. Height > 15m
B2	8	Emergent Trees. Height > 15m
B3	9	Emergent Trees. Height > 15m
B4	8	Emergent Trees. Height > 15m

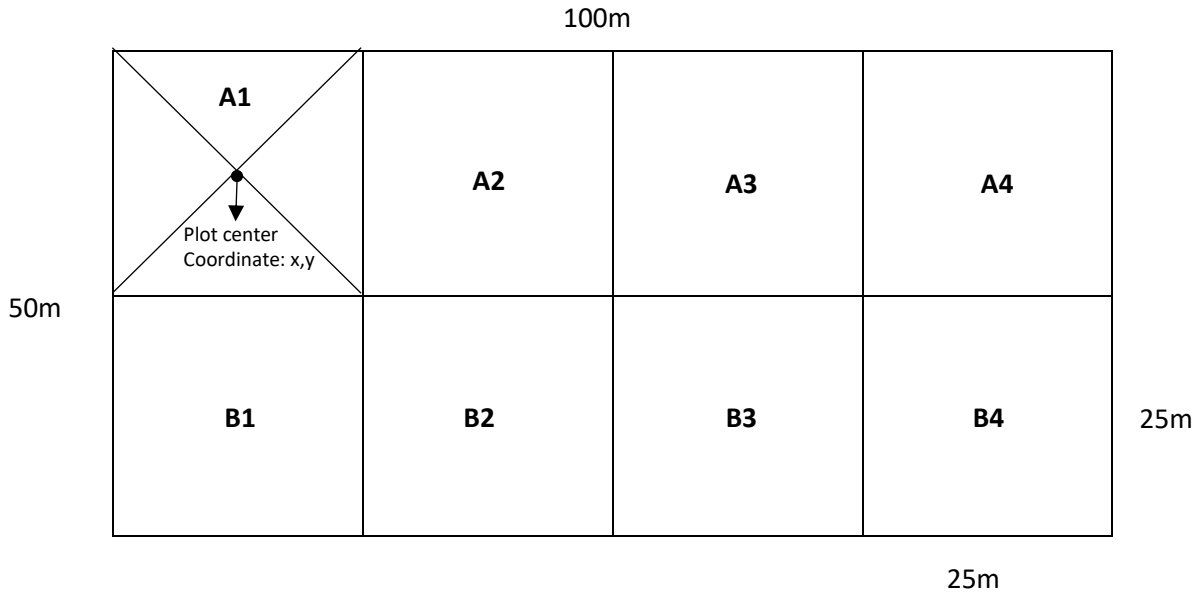


Figure 3.4. Dimensions of plots created within 0.5-ha plot

The current research was focused on quantifying biomass at tree level. A careful individual tree position measurement was therefore a key objective. Tree position was recorded using two different devices: a Trimble Laser Ace 1000 Range finder, which can offer up to 50 cm accuracy after post processing; and a Garmin e-Trex handheld

GPS with 3-5 meters accuracy. A total station was used to record position of the edge and plot centre. The stem positions were positioned relative to the given plot location with Trimble Ace Range finder. The relative planimetric accuracy positions varies from ±0.1 m to ±0.5 m. Tree physical measurement were recorded in the field based on the experimental plan designed by the author, recorded in fieldwork Form Sheet (Figure 3.5) by using appropriate fieldwork equipment (Table 3.4).

DATA COLLECTION FORM

Sample Plot ID:
Name of Recorder: Date:
Stratum ID: Time:

Bearing from the Road		Bearing for the 1st tree from the centre of the plot		Plot centre	
x		x		x	
y		y		y	
Angle		Angle			

Plot Radius: Altitude:
Crown Density (%)

No	Individual Tree Position		DBH (cm)	Bole Height (m)	Crown Diameter (m)	Height (m)	Species ID	Species Name	Remarks
	x	y							
1									
2									
3									
4									
5									
6									
7									
8									

Figure 3.5. Fieldwork form design used to record tree physical measurement

Table 3.4. List of equipment and function details used in the fieldwork campaign

NO	EQUIPMENT	Quantity	REMARK
1	Tape DBH	2	Measure DBH
2	Trimble Geo-7x (with Laser Ace Range finder)	1	Record tree position
3	GPS Handheld	2	Record tree position
4	GPS Trimble (MYRTKNET)	2	Plot sampling (survey)
5	GPS Topcon	2	Plot sampling (survey)
6	LAI Ceptometer/Hemispherical Camera	1	Measure Leaf Area Index (LAI) and forest cover
7	Vertex Hypsometer	1	Measure Tree Height
8	Total Station - Topcon -Piket	1	Plot sampling (survey)
9	Prism	2	Plot sampling
10	Tape	1	Measure Crown Diameter
11	Tag and tape	100 tagging	Tree Tagging
12	Monumentation -cement -nail -rope -iron pipe	4	Plot setup
13	Documentation -Paper		Record biometric data and physical measurement

3.7.3 Data Processing and Tree mapping

Data processing involved for fieldwork data was; (i) GPS control point coordinates using the Total Station, and (ii) re-projection of tree data and plot boundary through geographical mapping. The technique used to provide the GPS control point coordinates at PFR was using Real Time Kinematic Virtual Reference Station (RTK VRS) method. Two observations were made at two GPS control stations located at PFR. Real time accuracy at 95% confidence interval was reported with the accuracy of the observation on horizontal was within 4cm and on vertical was within 6cm. This range was acceptable to be used as control coordinates. The observation had been tied

to the two nearest Continuously Operating Reference Station (CORS) station, which are:

- KLAW-Klawang
- BAHA-Bahau

For orthometric height, WMGeoid04 was used to determine the height for each point set up and the coordinates of the control points was established between the two control points (Figure 3.6 and Figure 3.7). Meanwhile, all physical data was being gathered from all parties and converted to shapefiles format. The spatial data within 0.5 ha plot was projected to datum WGS84 UTM 48N across all data sets in order to match with the projection of the LiDAR data (Figure 3.8).

CONTROL POINT: UITM 01

Coordinate Projection: GDM2000

Latitude	2°58'06.60241"
Longitude	102°17'49.20535"
Elevation	105.789 m

Coordinate Projection: WGS84

Latitude	2°58'06.60218"
Longitude	102°17'49.20855"
Elevation	106.544 m

GRID GDM2000

N	28356.155
E	39721.419
Elevation	105.789 m

Figure 3.6. Control point 1

CONTROL POINT: UITM 03

Coordinate Projection: GDM2000

Latitude	2°58'06.90156"
Longitude	102°17'49.61355"
Elevation	106.974 m

Coordinate Projection: WGS84

Latitude	2°58'06.93261"
Longitude	102°17'49.60575"
Elivation	107.128 m

GRID GDM2000

N	28365.357
E	39733.907
Elevation	106.974 m

Figure 3.7. Control point 2

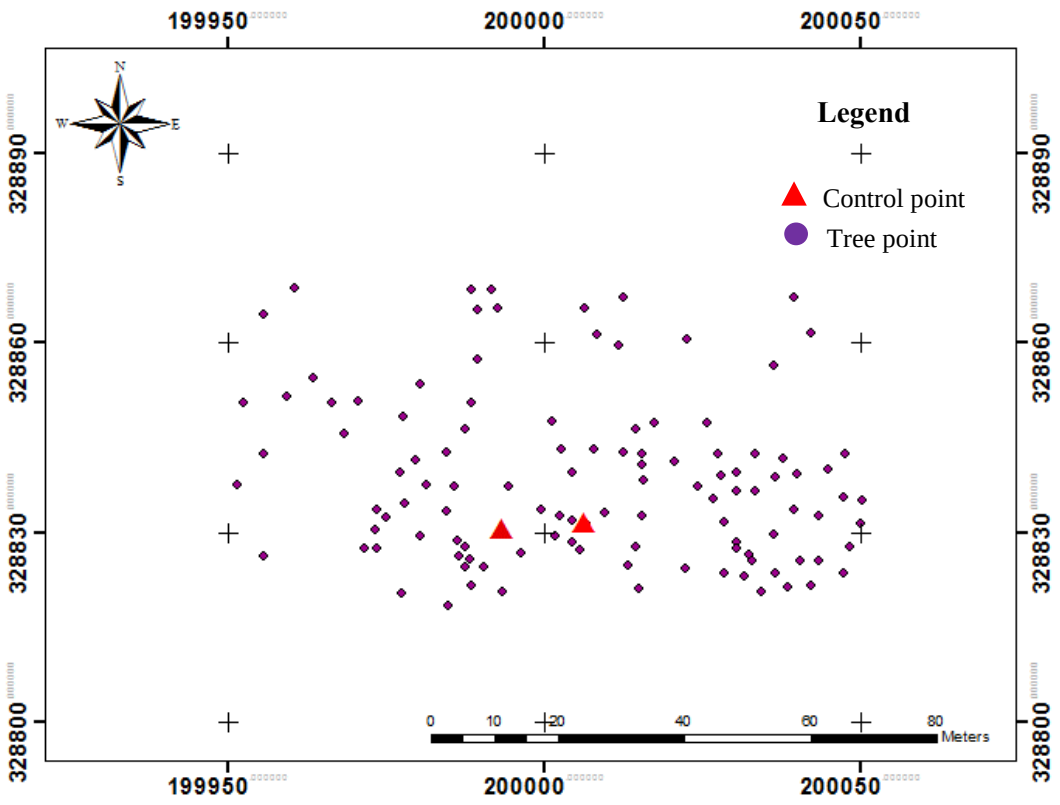


Figure 3.8. Location of individual tree measured across 0.5-ha plot at PFR

3.7.4 Description of Fieldwork Parameters

3.7.4.1 Diameter at Breast Height (DBH)

The tree DBH is one of the important variables used in the estimation of tree and stand volume. Generally, it is defined as the outside bark diameter at breast height and the breast height is defined as 1.37 meters (4.5 feet) above the forest floor on the uphill side of the tree. The tree structure in the study plot is clear straight with a gradually tapering trunk, and measuring the DBH was straightforward. In this fieldwork, a diameter tape is used for measuring tree diameter due to its convenient size and relative accuracy. All trees with the size of DBH ≥ 10 cm was taken for measurement in this study area. There are 51 trees from mixed species that have a DBH within the range from 10 cm – 20 cm, 22 trees within the DBH range from 20cm -30 cm, 11 trees within the DBH range from 30cm – 40 cm, 6 trees within DBH from 40 cm – 50 cm and 18 trees with DBH size > 50 cm (Figure 3.9). The same comparison was made to FRIM site (Figure 3.10).

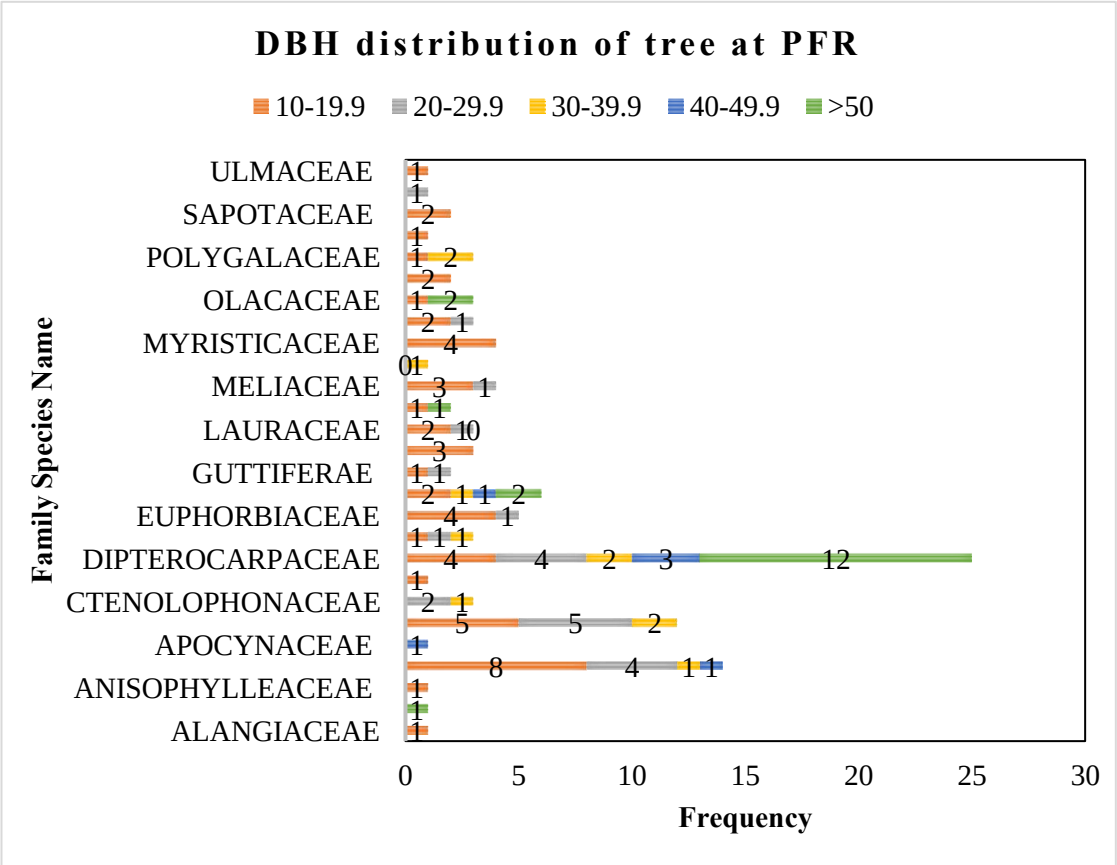


Figure 3.9 The DBH distribution of trees in 0.5-ha plot at PFR

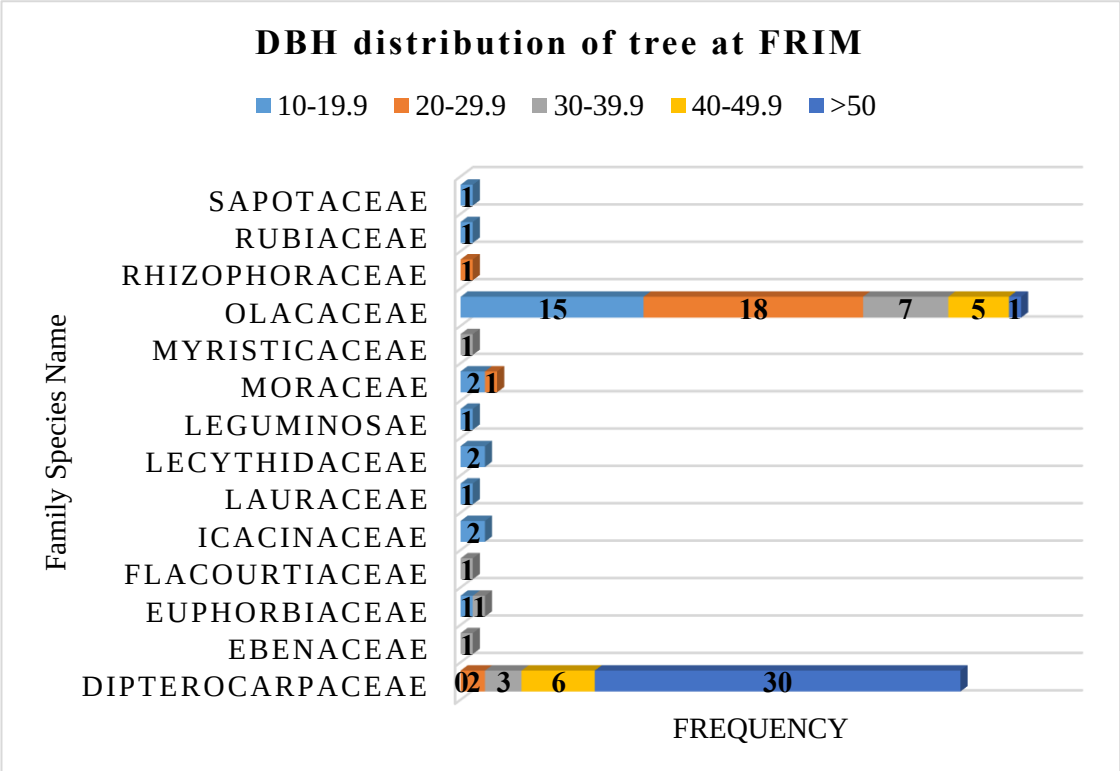


Figure 3.10 The DBH distribution of trees in 0.5-ha plot at FRIM

3.7.4.2 Total Height and Bole Height

Total height is one of the most fundamental measurements in forest inventory and is an important variable in the quantitative assessment of forest biomass, carbon stocks, growth and site productivity (Andersen *et al.* 2006). In the forestry context, total height is defined as the vertical distance between the ground level and tip of the tree (Husch *et al.*, 1972) and bole height may be defined as the distance from the base of the tree to the base of the first living branch that forms a part of the tree crown (Brack 1999). Measurement of total height and bole height are difficult for tall trees, especially in dense forest with trees close together and overlapping crowns. In this fieldwork, the total tree height (Figure 3.11 and Figure 3.12) and bole height (Figure 3.13 and Figure 3.14) are measured using a Vertex Hypsometer. The vertical accuracy is expected to be ± 1.0 m. On the first four plots at PFR site, starting with plot A1 until A4, a measurement was taken based on the aim of DBH size ≥ 10 cm. On the subsequent plot, starting with plot B1 until B4, based on the experience of doing fieldwork on the previous day, an extra criterion was added on selecting trees in the study plot which is aimed for emergent trees with height ≥ 15 meters. However, all tree recorded in this study less than 100 meters in height (Figure 3.15 and 3.16) and amount of AGB estimated almost at same figure (Table 3.5).

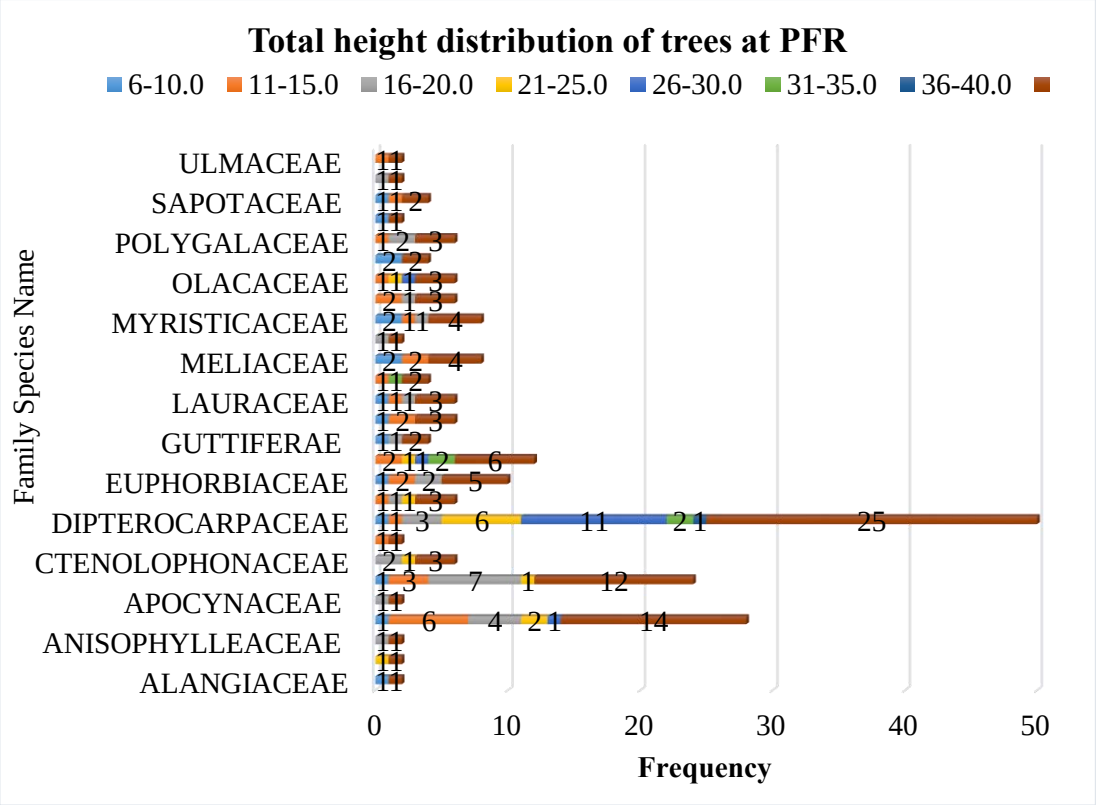


Figure 3.11. The total height distribution of trees in 0.5-ha plot at PFR

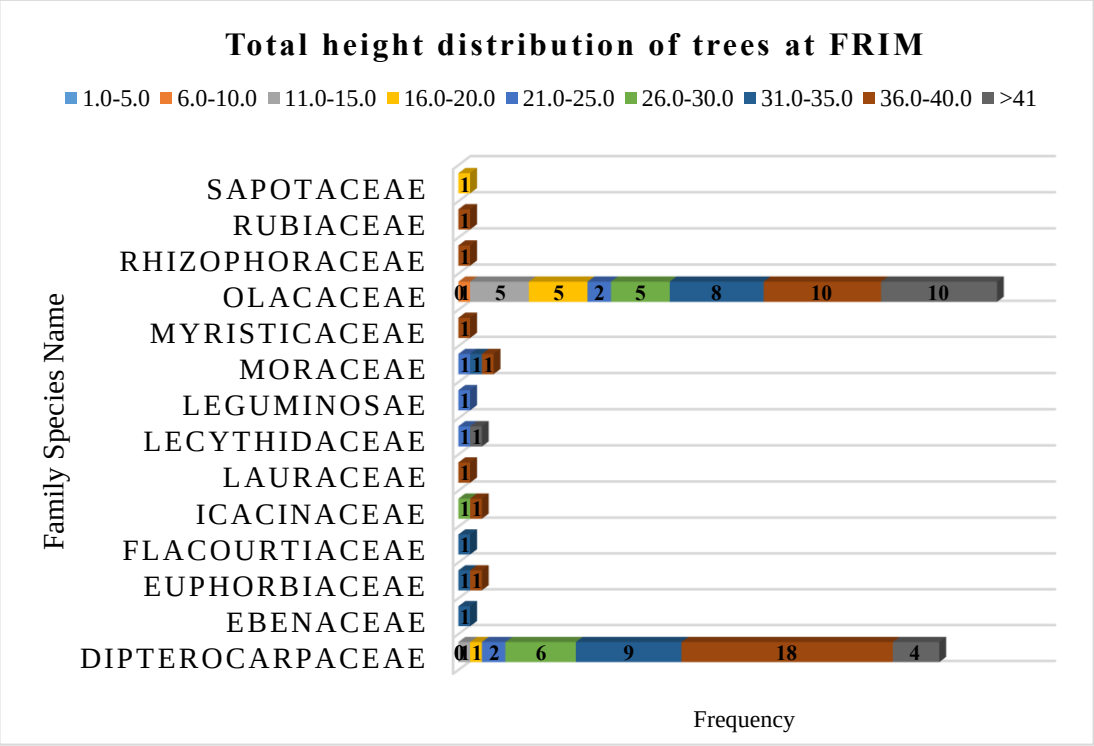


Figure 3.12. The total height distribution of trees in 0.5-ha plot at FRIM

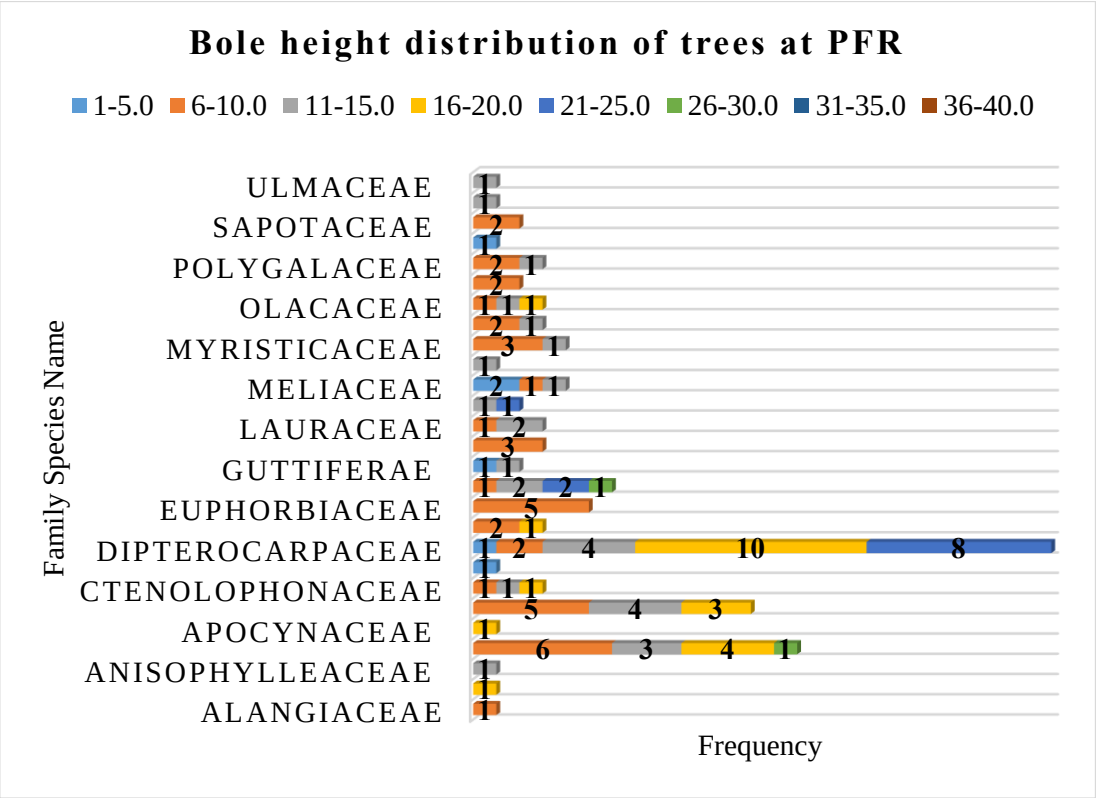


Figure 3.13. The bole height distribution of trees in 0.5-ha plot at PFR

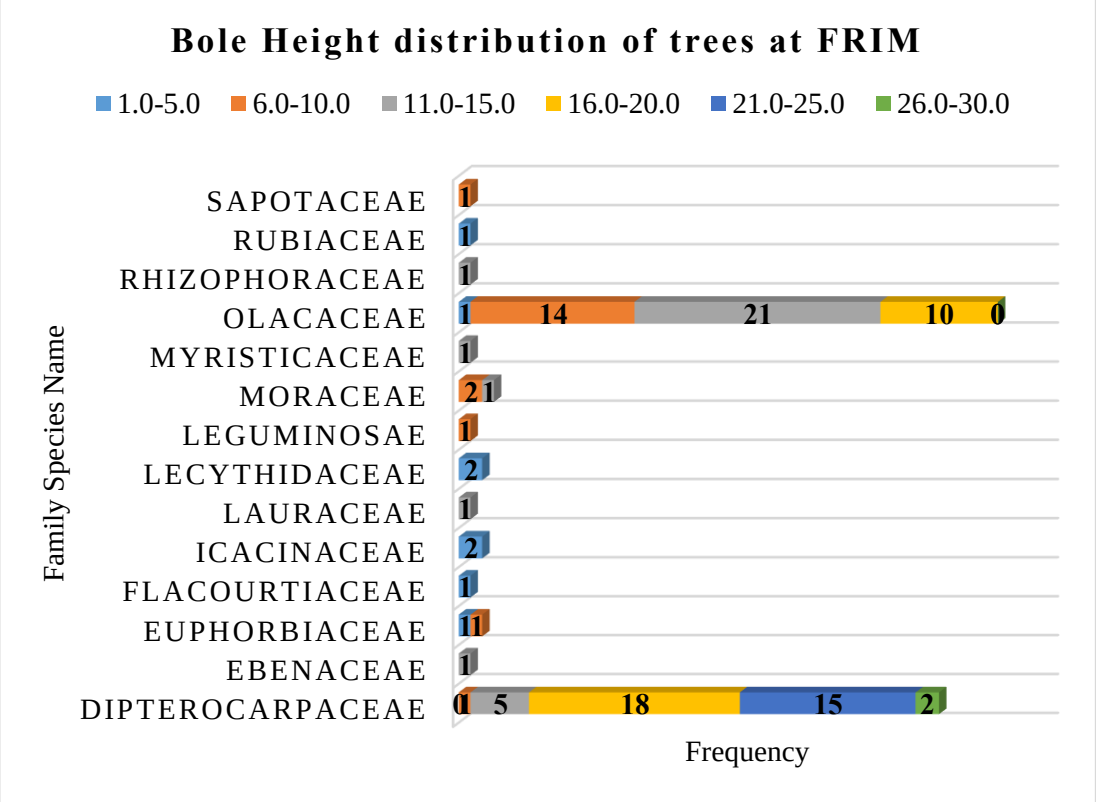


Figure 3.14. The bole height distribution of trees in 0.5-ha plot at FRIM

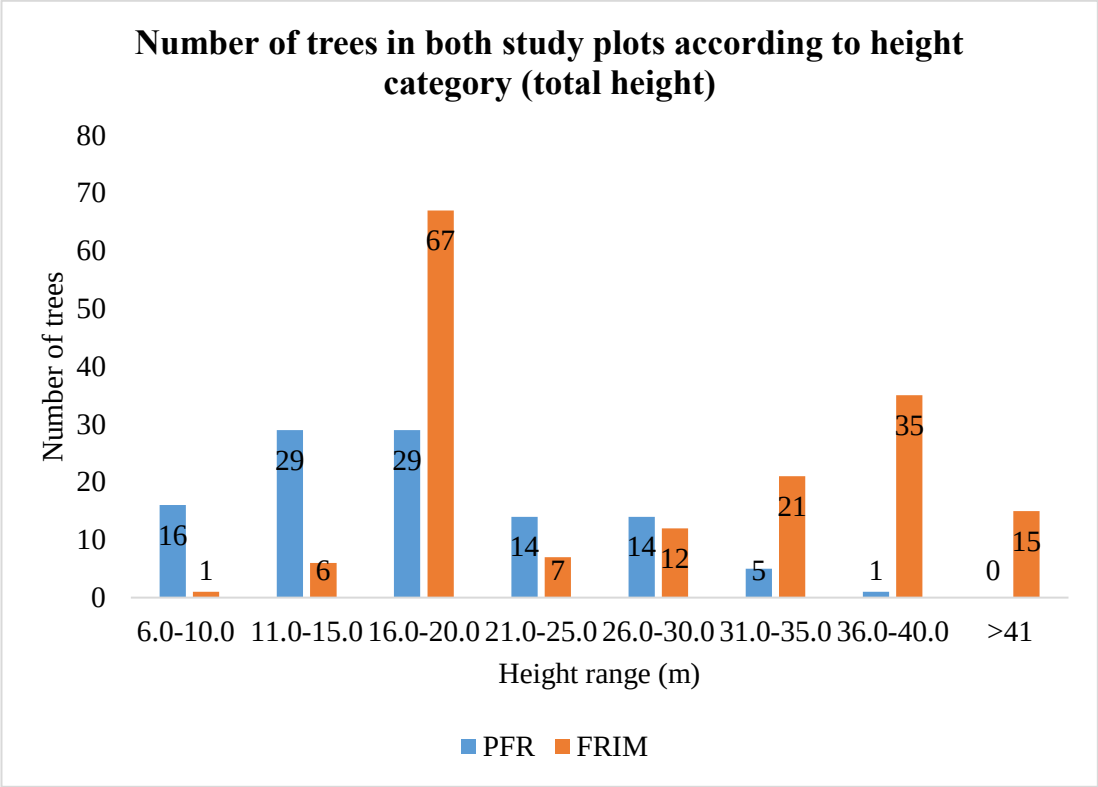


Figure 3.15 Numbers of trees in both study plot according to total height category

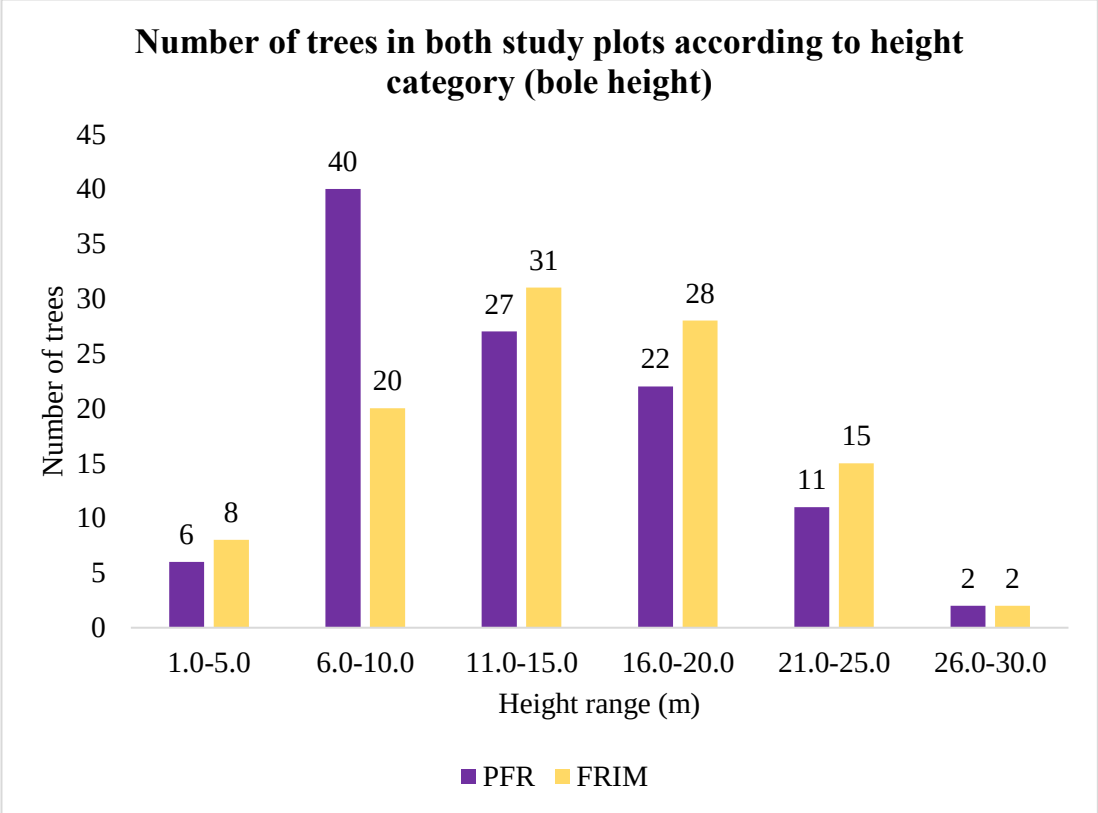


Figure 3.16 Numbers of trees in both study plot according to bole height category

Table 3.5. Characteristics of Individual tree field sample from both site

Forest Site	Parameters	Min	Max	Mean	Standard Deviation
PFR	Tree Height	7.30	39.90	18.16	6.87
FRIM	(m)	12.90	49.60	38.93	9.36
PFR	DBH (cm)	10.10	82.30	29.84	19.33
FRIM		10.00	82.20	37.92	20.86
PFR	Volume (m^3)	0.02	8.00	1.03	1.70
FRIM		0.12	9.14	2.46	2.47
PFR	AGB (kg)	8.70	8067.16	918.74	2219.91
FRIM		122.1	7771.90	2421.80	2162.92

Study by T.R Feldpausch *et al.* (2011), found tropical tree H: D allometry to be modulated by geographic location, environment and forest structure. A simple assessment has been made to examine the relationship between the forest structural parameters for PFR site. Power function, log (H): log (DBH) was chosen in fitting height to diameter relationships. A direct interpretation from power fitted function, shows that height was related to DBH with $R^2 = 0.7$ (Figure 3.17). Since AGB is a direct relationship with DBH (explained in detail in section 3.7.4), it can therefore be deduced that AGB is also related to DBH and height on individual tree and the AGB for the field data was calculated using the allometric equation, as described in 3.7.3 section. The inclusion of height in allometric equations greatly improves the accuracy of individual tree biomass estimation (Chave *et al.* 2005) and height should be included as a parameter in biomass estimates wherever possible (T.R Feldpausch *et al.* 2011).

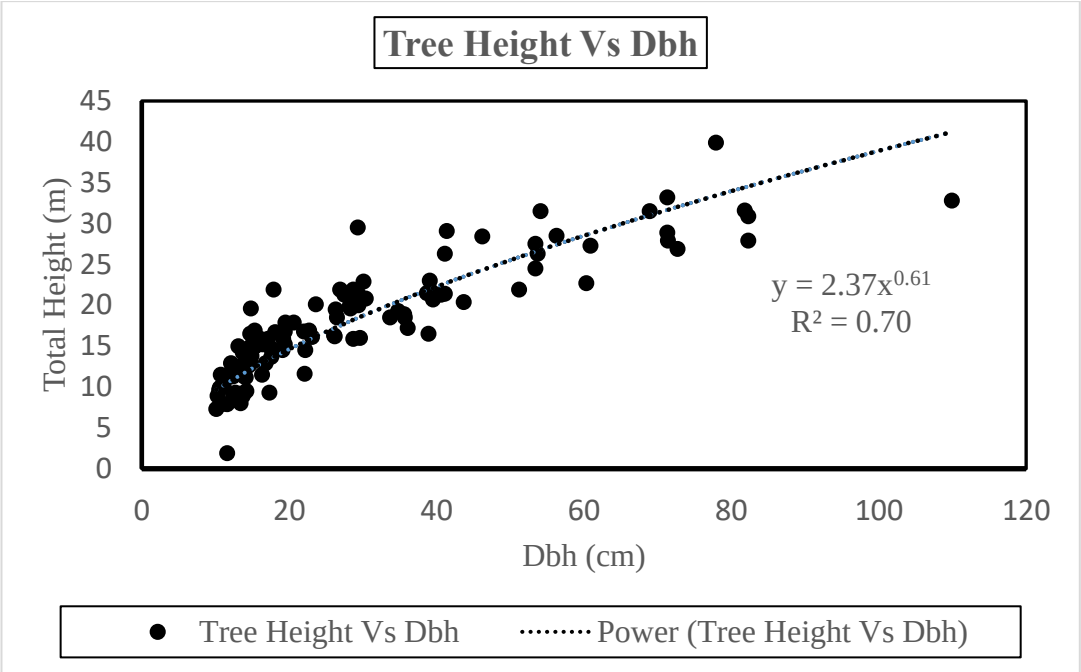


Figure 3.17. Scatterplots of tree height (in meter) against DBH (in centimetre) for all measured tree in the field from PFR site

3.8 Tree species

From the fieldwork, 142 individual trees were taken for measurement and 105 trees were used for final assessment at PFR site. While 104 trees at FRIM site were selected from the inventory sample. These trees came from 27 different family species (Table 3.6) and 25 different family species (Table 3.7) individually.

Table 3.6. Summary of the species information for tree measured in 0.5 ha plot PFR

No	Species Family	Species scientific name exist for tree measured in the field plot	Number of trees belong to this family
1	Alangiaceae	Alangium ebenaceum	1
2	Anacardiaceae	Pentaspadon Motleyi	1
No	**Species Family**	**Species scientific name exist for tree measured in the field plot**	**Number of trees belong to this family**
3	Anisophylleaceae	Anisophyllea	1
4	Annonaceae	Goniothalamus tortilipetalus, Polyalthia hypoleuca, Monocarpia Marginalis, Alphonsea maingayi, Uvaria foetida, Polyalthia jenkinsii, Polyalthia jenkinsii, Cyathocalyx pruniferus, Xylopia magna, Polyalthia Hypoleuca, Anicosanthum Fuscum	14
5	Apocynaceae	Dyera Costulata	1
6	Burseraceae	Santiria laevigata, Dacryodes costata, Santiria apiculata, Santiria Rubiginosa, Santiria Laevigata, Canarium Littorale, Dacryodes Laxa	12
7	Ctenolophonaceae	Ctenolophon Parvifolius	3
8	Dilleniaceae	Dillenia Sumatrana	1
9	Dipterocarpaceae	Shorea multiflora, Shorea macroptera, Dipterocarpus sublamellatus, Hopea mengarawan, Shorea leprosula, Dipterocarpus cornutus, Shorea acuminata, Anisoptera Megistocarpa, Vatica Bella,	25
10	Ebenaceae	Diospyros kaki, Diospyros buxifolia,	3
11	Euphorbiaceae	Neoscortechinia kingii, Blumeodendron tokbrai, Drypetes kikir, Cleidion spiciflorum,	5

		Mallotus Griffithianus, Aporosa Prainiana, Aporosa Bracteosa	
12	Fabaceae	Archidendron bubalinum, Sindora echinocalyx, Callerya atropurpurea, Koompassia malaccensis	6
13	Guttiferae	Garcinia eugeniifolia	2
14	Lamiaceae	Teijsmanniodendron coriaceum, Alangium ebenaceum	3
15	Lauraceae	Litsea ferruginea, Litsea Magnifica Cryptocarya rugulosa, Phoebe Grandis	3
No	**Species Family**	**Species scientific name exist for tree measured in the field plot**	**Number of trees belong to this family**
16	Leguminosae	Archidendron microcarpum, Koompassia Malaccensis	2
17	Meliaceae	Reinwardteodendron cinereum, Chisocheton erythrocarpus, Aphanamixis Polystachya, Chisocheton Erythrocarpus	4
18	Moraceae	Artocarpus lanceifolius, Artocarpus Elasticus,	1
19	Myristicaceae	Knema kunstleri, Myristica cinnamomea, Knema scortechinii	4
20	Myrtaceae	Syzygium duthianeum, Syzygium castaneum, Syzygium gratum, Syzygium griffithii, Syzygium Chloranthum	3
21	Olacaceae	Ochanostachys amentacea, Anacolosa frutescens	3
22	Phyllanthaceae	Aporosa prainiana, Aporosa bracteosa, Cleistanthus Maingayi	2
23	Polygalaceae	Xanthophyllum stipitatum, Xantophyllum rufum	3
24	Sapindaceae	Nephelium hamulanthum	1
25	Sapotaceae	Madhuca pasohensis	2
26	Trigoniaceae	Trigoniastrum Hypoleucum	1
27	Ulmaceae	Gironniera parvifolia, Gironniera Nervosa	1

Table 3.7. Summary of the species information for tree measured in 0.5 ha plot FRIM

No	Species Family	Species scientific name exist for tree measured in the field plot	Number of trees belong to this family
1	Dipterocarpaceae	Dipterocarpus baudii, shorea macroptera Shorea bracteolata	41
2	Ebenaceae	Diospyros wallichii	1
3	Euphorbiaceae	Elateriospermum tapos	2
4	Flacourtiaceae	Paropsia vareciformis	1
5	Icacinaceae	Gomphandra quadrifida	2
6	Lauraceae	Actinodaphne pruinosa	1
7	Lecythidaceae	Barringtonia fusiformis	2
8	Leguminosae	Archidendron bubalinum	1
10	Moraceae	Streblus elongatus Streblus elongatus Artocarpus elasticus	3
11	Myristicaceae	Myristica sp.	1
12	Olacaceae	Ochanostachys amentacea	46
22	Rhizophoraceae	Pellacalyx saccardianus	1
23	Rubiaceae	Timonus wallichianus	1
25	Sapotaceae	Palaquium gutta	1

3.8.1 Biomass Measurement – Allometric Equations

There is now an abundance of allometric equations developed for different forest scenarios. Research has shown that the tree allometric data of tropical trees varies with the type of forest and also depends on the height and diameter at different geographic

locations. The selection of the best allometric equation of AGB depends on the characteristics and composition of the forest study area. In this research, 3 different allometric equations were compared (Table 3.8). These are the ones developed by Kato *et al.* (1978), Basuki *et al.* (2009) and the pan-tropic equations by Chave *et al.* (2014). These allometric models is an allometric equations developed specifically for tropical rainforest and established with different tree parameters as independent variables. Each one has important criteria for the study area, which can be summarised as follows:

- Kato *et al.* (1978): The allometric model was developed using tree sampling at PFR 6-ha International Biological Program (IBP) plot which where the fieldwork was run; this is a site-species specific allometric equation. The tree allometric equations are developed by establishing the relationship between DBH and Height (D-H).
- Chave *et al.* (2014): The most recent published Pan-tropic multi-species allometric equation based on a huge dataset, where there are about 4,004 trees (diameter more than 5cm) from 58 sites all over the world including PFR, Negeri Sembilan. The equations are developed by establishing the relationship between DBH, Height and wood density (D-H-W).
- Basuki *et al.* (2009): This allometric equation was developed specifically for lowland dipterocarp forest in a secondary forest in Kalimantan, Indonesia. This equation is a species-specific allometric equation and was developed by establishing the relationship with DBH only (D).

The results will be elaborated in the following sections.

Table 3.8. Allometric functions developed for study site and from various study site for estimating above ground biomass

No.	Source	Allometric functions	Site
1.	Kato et al (1978)	$TAGB = W_s + W_b + W_l$ From the values of DBH and height of a tree, the TAGB was estimated by the summation of stem biomass (W_s), branch biomass (W_b) and leaf biomass (W_l) $W_s = 0.0313 (D^2H)^{0.9733}$ $W_b = 0.039 (D^2H)^{1.041}$ $1/W_1 = 1/(0.124 W_s^{0.794}) + 1/125$	Primary forest, peninsular Malaysia. PFR International Biological Program (IBP) plot. Plot taken in this research was conducted within this IBP plot.
2.	Chave *et al.* (2014)	$AGB = 0.0673 \times (\rho D^2H)^{0.976}$ ρ = wood density (g/cm^3)	PFR 50-ha plots and other Center for Tropical Forest Science (CTFS) plots.
3.	Basuki *et al.* (2009)	$\ln(TAGB) = 2.196\ln(D) - 1.201$ D = diameter at breast height (cm)	Secondary forest, Kalimantan, Indonesia

3.8.2 Above Ground Biomass and Carbon Stocks

There are two different approaches to biomass studies for tropical rainforest; (i) site-specific equations only are considered so as to achieve accurate estimation of above-ground biomass (Basuki *et al.* 2009, Kenzo *et al.* 2009); and (ii) local species-specific allometric equations are replaced by generalized allometric relationships, in order to reduce the need for intensive field measurements at every site (Chave *et al.* 2005). For instance, in 1 ha of tropical forest there may be as many as 300 different tree species, so using species-specific regression models is far more elaborate than in forest within

the temperate zone. Mixed species tree biomass regression models are usually based on a small number of directly harvested trees and include very few large diameter trees, thus not well representing the forest at large. In this context, the allometric equations chosen are based on generalized pan-tropic allometric equations. Basuki *et al.* (2009) only takes DBH to predict the biomass, Kato *et al.* (1978) takes DBH and Height only to predict the biomass while Chave *et al.* (2014) put all the important tree parameters like DBH, Height and wood density into the model. Trees come in different shapes and different weight, the correlation with DBH only doesn't allow for that. The availability of tree height in allometric model usually yields less biased estimates compare to allometric model that does not include height factor (Chave *et al.* 2014). Tree height has often been ignored in carbon-accounting programs because measuring tree height accurately is difficult in closed-canopy forests, but the increasing availability of LiDAR data is making tree height data more prevalent and the need and so there is a growing need for allometry that includes it as a parameter. Better calibration and analysis of tropical tree allometric equations are needed to avoid mismatches of otherwise convergent studies, whether from plot inventory or plot-inventory-calibrated remote sensing. Allometry equation by Basuki *et al.* (2009) are lacking in terms of this relationship.

The role of wood density in the allometric equation is more prominent for the mixed species than in genera (genus). The study site is abundant with different species. Wood specific gravity is an important predictor in AGB especially when a broad range of vegetation types is considered. From the three different allometry equations discussed here, only the allometry model developed by Chave *et al.* (2014) applies wood density

to the model. The two other models are developed based on DBH and height only. Based on the fieldwork, DBH, tree height and species information had been measured and identified and Chave equation is a tree multiple-parameter equation because it includes all of this, as it contains the shape of tree. Additionally, the model by Chave has been improved by including twice the number of trees from the previous study in 2005 and updated to the most recent years in 2014.

The importance of including wood density in biomass estimation can be examined for big trees, such as *Dipterocarpus cornutus* with a diameter of 77.9 cm has a dry weight of TAGB 8067 kg whereas *Dipterocarpus submellatus* with a diameter of 82.3 cm has a lower dry weight of TAGB that is 6053 kg. It is likely that the differences in wood density and tree architecture explain the differences in the dry weight of these two species. Trees of the same diameter, height and wood density can display a range of biomass value. The characteristics of these species show that their wood density, DBH and Height are: 0.66 g/cm^3, 77.9cm and 39.9m for *Dipterocarpus cornutus* and for *Dipeterocarpus submellatus* they are 0.63g/cm^3, 82.3cm and 27.9m respectively. In this research, the genus of Dipterocarp consists of several big trees, however adding wood density to the model does not significantly influence the β coefficient. Biomass results for individual tree measured at both site is listed in Appendix 2 and Appendix 3.

3.8.3 Carbon Stocks

Carbon stock of the tree was calculated from AGB using conversion factor 0.47 (IPCC, 2003). The results of carbon stocks calculated from different allometric equations was listed in Appendix A.

$$\textit{Carbon Stock} = 0.47 \times \text{AGB}. \quad (3.1)$$

3.9 Earth Observation Data

The data used in this research are discrete return airborne LiDAR. Digital Orthophoto was only used in this study for the purpose of point identifications. The details of the data are discussed in the next section.

3.9.1 Discrete return Airborne LiDAR

The LiDAR data for PFR obtained from a private Malaysian airborne company - using an IGI LiteMapper-5600 system with a Riegl Q560 LiDAR sensor scanning at a ± 22.5^0 at a line rate of 60 line/s. Data processing steps include the production of radiometrically calibrated data (level 1), traceable to national standards for derived geophysical data products (level 2), which followed the application of an atmospheric correction. Finally, data were geometrically rectified to the local geo-reference co-ordinate system with user-defined Ground Control Points (level 3). The data supplied were checked for quality and delivered as classified and unclassified point clouds in both ASCII XYZ and LAS formats with a projection of UTM_Zone_48N. The data have been validated and quality checked and any possible low points have been removed. Varying types of topography (mountainous, rolling or flat terrain) may affect the accuracy at which the elevation surface can be modelled. In such situations, it may be preferable to specify different accuracy requirements for the various terrain types and to design separate tests for each (ASPRS LiDAR Committee, 2004). Root Mean Square Error (RMSE) is utilized to assessed the vertical accuracy of dataset. The

LiDAR data received from the vendor recorded RMSE achieved for specific Land Class of Forest was 0.092 meter and proved that the accuracy of LiDAR data is within tolerance of 0.15 m in vertical offset. The average point density was 8.8 points per square meter.

The LiDAR data from FRIM site was collected from FRIM Geospatial programme department using an ALTM Gemini laser system. The data supplied at the same standard as the data received from PFR with the average point density 7.5 points per square meter (Table 3.10).

Table 3.10. LiDAR acquisition parameters

LiDAR sensor	**IGI Lite Mapper-5600 Riegl Q560 (PFR)**	**ALTM Gemini laser system (FRIM)**
Pulse Rate	Range between 70 KHz to 240 KHz	70 KHz
Scan Angle	$\pm 22.5^0$	$\pm 25^0$, increments of $\pm 1^0$
Scan Pattern	Regular	Regular
Effective rate	46,667 Hz	33- 167 KHz
Line/sec	Max 160	Max 160
Flying height	700-1000m	150- 4000 m nominal
Laser points/m^2	8.8 points/m^2	7.5 points/m^2
Max Above Ground Level	1040m (3411ft)	5000 m (16404ft)
Data Format	ASCII XYZ and LiDAR exchange format (LAS)	LiDAR exchange format (LAS)

3.10 Tree Positioning

In a post-processing step, tree location was manually checked and co-registered to LiDAR Canopy Height Model (CHM) (elaborate in detail in Chapter 4). To obtain an interpretable best fit of the tree pattern with CHM, the tree pattern was visualized and manually moved in ArcMap 10.1. The information from CHM shows the local object

heights and can be derived by subtracting the Digital Terrain Model (DTM) from a Digital Surface Model (DSM). The stem location was only manually corrected if the initial position from the field survey showed gross error. After manual co-registration, the estimated absolute planimetric accuracy of the stem location is ±0.5 m. This step was explained in detail in Chapter 4

3.11 Fieldwork Challenges and Chapter Summary

The initial fieldwork was run in October 2014 at PFR in conjunction with Natural Environment Research Council (NERC) and The Airborne Research & Survey Facility (ARSF) flight campaign to Malaysia. However, due to poor weather and restricted timing window, the NERC airborne campaign did not manage to collect the LiDAR and Hyperspectral data as initially planned for this thesis (which would have focused on full waveform LiDAR and Hyperspectral data for quantification of aboveground biomass). As an alternative, the discrete return LiDAR data that was obtained were used to carry on the aims of the dissertation without substantially changing the initial aim.

CHAPTER 4

Modelling Individual Tree Aboveground Biomass Using Discrete Return LiDAR in Lowland Dipterocarp Forest of Malaysia

A portion of the research in this chapter is accepted in the following journal paper:
Wan-Mohd-Jaafar W.S., Woodhouse I.H., Silva C.A., Omar H., Hudak A.T, (2017). Modelling individual tree aboveground biomass using discrete return LiDAR in lowland dipterocarp forest of Malaysia. *Journal of Tropical Forest Science, 29(4) p 465-484*

My contributions in this research include collected, processed and analysed LiDAR data, interpreted results, prepared the manuscript, and coordinated revisions of the manuscript. Iain Woodhouse supervising the work, assisted with interpretation of the results and reviewed the manuscript. Carlos Silva, assisted with interpretation of the results, graphics improvement and reviewed the manuscript, Hamdan Omar, assisted in designing the fieldwork experiment, assisted in data collection, and provide access to compilation of primary and secondary dataset and reviewed the manuscript and Andrew Hudak reviewed the manuscript

4.1 Introduction

Tropical forests in Southeast Asia have declined acutely over the past several decades (Laurance 2007). In particular, according to a new global forest map in partnership with Google, Malaysia had the world's highest rate of forest loss between 2000 and 2012 (Butler 2013). Reducing Emissions from Deforestation and forest Degradation (REDD+) is the international framework for conserving and enhancing carbon stocks of forested area in the tropics (UNFCCC 2007). For REDD+ implementation, accurate estimation and monitoring of carbon stocks are required at the national and subnational levels. To establish robust and transparent monitoring systems, a combination of ground-based sampling and remote sensing approaches was recommended (UNFCCC 2009). Aboveground biomass (AGB) of trees in tropical forests account for a significant part of the total carbon pool (Houghton et al. 2001). Therefore, estimating AGB is critical to accurately quantifying carbon stocks in the tropics (Gibbs et al. 2007).

Tropical forests are known for their complex stand structure and abundant diversity in species composition (Steininger 2000) and estimating AGB from remote sensing data in this dense forest is challenging. Satellite-mounted optical sensors have been widely used to estimate AGB (Anaya et al. 2009). However, optical sensors acquire information from the upper canopy and are unable to measure the three-dimensional structure, including canopy height and sub-canopy topography (Lu 2006), which limits their utility to quantify AGB in tropical forests with complex canopy structures. Radar sensors (e.g., ALOS/PALSAR) use active microwave signals to generate an image,

and these can be used to determine forest vertical structure (Gibbs et al. 2007), even in areas of high cloud cover such as the tropics. Radar sensors can be used for relatively young or homogeneous forests, but their accuracy and sensitivity decrease in old-growth forests unless longer wavelengths are used (Hamdan et al. 2015).

Light detection and ranging (LiDAR) emits laser pulses and measures the return time of scattered returns to directly estimate the height and vertical structure of forests (Dubayah & Drake 2000; Lefsky et al. 2002). LiDAR on airborne platforms can be acquired at high sampling density with excellent geometric accuracy and can reveal AGB variation at fine spatial scales (Reutebuch et al. 2005; Mallet & Bretar 2009). LiDAR is therefore well placed to bridge the scale gap between satellite observations and field measurements (Asner 2009).

LiDAR remote sensing systems can be distinguished based on the way in which return signals are recorded (discrete return or waveform), scanning pattern (profiling or scanning), platforms (airborne, spaceborne, or ground based) and footprint sizes. The most common configuration of LiDAR systems is airborne, small footprint, discrete return scanning LiDAR, as used in this study. Airborne discrete return LiDAR has been used in a large number of studies for mapping biomass mainly using two kinds of approaches: (i) area-based and (ii) individual tree-based methods.

In this study, the individual tree-based method was considered at Pasoh Forest Reserve in Peninsular Malaysia. This site contains mixed species and is dominated by trees from the *Dipterocarpaceae* family, which is common in lowland dipterocarp forest.

Lowland dipterocarp forest is one of the most species-rich communities in the world, with more than 200 species per hectare (Symington 1943; Wyatt-Smith 1964). Individual tree detection is seen as the most relevant approach to extract tree structural attributes in tropical rainforest characterised by a complex three-dimensional structure. LiDAR forest inventory methodologies based on individual tree detection have been widely studied, but are not widely used in practice, due to the difficulties of tree detection in various forest conditions, especially in dense, closed-canopy tropical forests (Kaartinen et al. 2012).

The most widely used LiDAR metrics for AGB prediction are various height metrics that are associated with field measurements through empirical models (Kaartinen et al. 2012). LiDAR metrics can be calculated based on first return, last return or all of the returns (Chen 2013). In this study, all returns were used to maximize the information content. Unlike most of the published algorithms that detect individual trees from a LiDAR-derived raster Canopy Height Model (CHM), this study worked directly with the LiDAR point cloud data combined with field data to distinguish individual trees and to estimate individual tree metrics. The CHM is a raster image interpolated from LiDAR points depicting the height from the ground to the top of the vegetation canopy (Khosravipour et al. 2014). As a result, the CHM can have inherent errors and uncertainties from a number of sources (Khosravipour et al. 2014). However, by directly interpolating the raw LiDAR point cloud to extract individual trees, the measurements are not affected by the errors associated with interpolation, and the important 3D forest parameters can be extracted directly from the LiDAR returns that make up each tree (Li et al. 2012).

There are limited studies published on the use of LiDAR and the applications of LiDAR data for estimating AGB in Malaysia (Ismail & Manaf 2011). The closest studies related to AGB and airborne LiDAR was conducted by Hamdan et al. 2015, but this is not an individual tree based approach and does not extracted tree parameters such as height, base height and crown diameter from the LiDAR point clouds, instead based on LiDAR-derived canopy height model (CHM). To the author's knowledge, the use of direct LiDAR point cloud to detect individual trees and extraction of LiDAR height metrics in tropical rainforest of South East Asia has been little studied and this is one of the first studies to implement this approach for individual tree LiDAR-AGB modelling in Malaysia.Due to the structural complexity of tropical rainforests, further research is needed to identify the relationship between AGB measured in the field and LiDAR height metrics, and to determine how these relationships impact the accuracy of predictive models. This research integrates, tree-level field-sample data with LiDAR variables to predict AGB in tropical rainforest. The goal of this study was to model individual tree AGB based on trees that were mapped in the field, with the intention that the model could later be applied to a wider area. The immediate objectives were to: (1) develop a non-linear AGB model based on field sample plot and LiDAR data, and (2) validate the model in terms of accuracy and precision.

4.2 Materials and Methods

4.2.1 Field Measurement and field data collection

The data used in this study include a vegetation field sample data collected in 2014 and LiDAR data collected in 2012 in PFR, as well as the vegetation field sample data and LiDAR data in FRIM both collected at the same year in 2014.

Fieldwork was conducted at PFR, Negeri Sembilan, Malaysia from 28th October to 30th October 2014. Six main forest parameters were collected; horizontal position (x, y) of individual trees, diameter at breast height (DBH), total height, bole height, crown diameter and tree species. Stratified random sampling of a rectangular area of 50 m x 100 m (0.5 ha) was split into 8 experimental plots with dimension of 25 m x 25 m (Figure 4.1). All trees with DBH >10cm were measured within plots A1, A2, A3 and A4. The criteria were changed for plot B1 to B4 such that in these plots only emergent trees with height > 20 meter were measured. The measurement strategy was designed in such a way as to facilitate the mobility of sampling work in the field. 105 individual trees were used for final assessment.

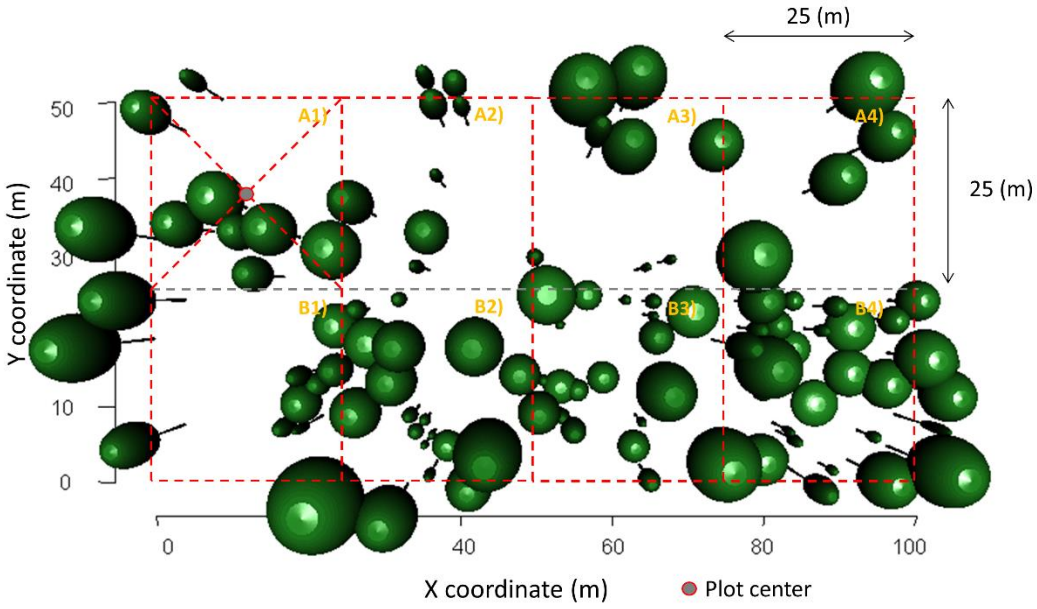


Figure 4.1. 3D representation of LiDAR dataset for plots created within 0.5-ha at Pasoh Forest Reserve. Ellipsoids have been used to represent the data collected in the field.

Considering that the distribution of the AGB sample data deviates far from a normal distribution, the sample size of the selected 105 trees is possibly too small to build a reliable regression model representative of all species across both study areas. To address this issue, the field sample acquired in early 2014 from FRIM were evaluated to determine whether it could be incorporated into the regression model. Since both PFR and FRIM are under the authority of FRIM, a similar sampling method and plant measurement protocol was applied at FRIM as at PFR. FRIM is similar to PFR in terms of topography, vegetation composition and structure and land use history, because both forest are managed under the authority of FRIM. Some environmental and successional differences existed between PFR and FRIM. However, the AGB characteristics are likely to have more important impact on the LiDAR metrics than the environmental factors (assuming a relationship exists between the AGB and the LiDAR variables). To test this assumption, an analysis of covariance (ANCOVA) was conducted in this study as detailed in section 4.4.1

The trees locations were determined using the geographic coordinates of the plot centres and the direction and distance of trees relative to the plot centre. The plot centres were measured with a handheld global positioning system (GPS) device and the locations were post processed with local base station data, resulting in an average error of approximately 0.5 m horizontally. Tree heights were measured using a hypsometer and the DBHs and crown diameter were measured using a diameter tape (d-tape). Crown diameter was measured in four cardinal directions with respect to tree trunk. Statistics describing the trees are in Table 4.1.

Table 4.1. Characteristics of Individual tree field sample from both site

Forest Site	Parameters	Min	Max	Mean	Standard Deviation
PFR	Tree Height	7.30	39.90	18.16	6.87
FRIM	(m)	12.90	49.60	38.93	9.36
PFR	DBH (cm)	10.10	82.30	29.84	19.33
FRIM		10.00	82.20	37.92	20.86
PFR	Volume (m^3)	0.02	8.00	1.03	1.70
FRIM		0.12	9.14	2.46	2.47
PFR	AGB (kg)	8.70	8067	919	2220
FRIM		122.1	7772	2422	2163

4.2.2 Above Ground Biomass (AGB) estimation

An abundance of allometric models to calculate AGB have been developed for South East Asian tropical secondary and Dipterocarp forests e.g. (Basuki et al. 2009, Niiyama et al. 2010). The selection of allometric equation for AGB are dependent on the characteristics and composition of the study area. In this study allometric equations for calculating AGB from field measurement were selected from Chave et al. 2014. This is a pan-tropical multispecies allometric equation whereby the study site is one of the many test sites the Center for Tropical Forest Science (CTFS) used to develop the allometry model. Moreover, the chosen equation (Chave et al. 2014) is a tree parameter equation incorporating DBH, height and wood density as predictor variables, describing the shape of the tree. Wood density is an important determinant of AGB, especially when a broad range of vegetation types is considered, and this model has been improved by including twice the number of trees as a previous study in 2005 and updated to the most recent years in 2014. A careful selection of allometric equations is important to reduce the uncertainty in estimating AGB. The equation used is as follows;

$$AGB = 0.0673 \times (\rho D^2 H)^{0.976} \quad . \qquad (1)$$

Where ρ is wood density (g/cm^3), D is DBH (cm), H is tree height (m)

4.3 LiDAR Operations

The details of LiDAR data specification were explained in previous chapter in section 3.9.1.

4.3.1 Data Post-processing

FUSION software was used to process the LiDAR data to generate three main products: the digital terrain model (DTM), the digital surface model (DSM) and the canopy height model (CHM) for use in this study. FUSION software system was developed by McGaughey (2009) provides visualization and analysis capabilities for LiDAR projects that can handle large data sets for free. The *catalog* function was used to evaluate the LiDAR characteristics.

A Digital Terrain Model (DTM) was created in two steps from the discrete return LiDAR data: first, the data were filtered to remove the above-ground returns using algorithm an adapted from Kraus & Pfeifer (1998); this was performed by using the *groundfilter* function, second the DTM was created by calculating the average elevation from the remaining (ground) LiDAR returns within a cell (cells that contain no points were filled by interpolation using neighbouring cells) by using function *gridsurfacecreate*. This algorithm was developed for vegetation-covered areas based on robust linear prediction (Kraus & Mikhail 1972), which has been widely accepted

by researchers. The digital surface model (DSM) which represents the Earth's surface and includes the trees and other objects on it, was created using *canopymodel*. The discrete return point clouds were then normalized against the ground surface height and extracted for each plot using the coordinates of the lower left and upper right plot corners by using the *clipdata* function. After heights were normalized, the canopy height model (CHM), which represents the height of the forest was generated by using *canopymodel* function. The CHM was created for visualization and image interpretation purposes and for manual co-registration between field sample data and LiDAR data, but not for LiDAR metrics extraction. Rather, individual tree LiDAR metrics were extracted directly from the normalized point cloud data. All point cloud data processing was performed using FUSION software (McGaughey 2014).

4.3.2 Individual tree extraction: - Co-registering LiDAR and Field Sample data

The development of the co-registration procedure was based on field sample data collected during fieldwork. The attributes relevant for this study are given in Table 3. The positions of the plot edges were georeferenced with a total station. Coordinates of each tree were determined by manually seeking the optimum fit between tree positions and heights measured by forest sampling and CHM. To do this, the absolute positions of the trees within each plot were calculated from the geographical coordinates of the sample plot centres and the coordinates of the individual trees measured from field. To obtain an interpretable best fit of the tree pattern, these coordinates were then converted into ArcGIS shapefiles, which, in combination with the field height of each tree and crown diameter measured in the field, formed a polygon representing the

crown dimensions of each tree. Each of the tree crown polygons was assigned a unique identification number (ID) and projected to the same projection as the LiDAR (UTM Zone 48N).

To facilitate visual comparison with the LiDAR Canopy Height Model (CHM), the field tree crown polygon was visualized and manually moved to best fit the shape and height of the CHM. Errors of this manual co-registration method are expected to lie at the subpixel level (i.e. < 1.0 m). Out of the 142 individual trees, 3 trees were initially removed due to the condition of being broken at first branch, which left 139 individual trees to be assessed manually. From these, 105 could be unambiguously manually co-registered. After manual co-registration, the estimated absolute planimetric accuracy of the tree location was ± 0.5 m. After carefully co-registering the individual tree polygons with the LiDAR data, using the X, Y coordinates and the crown diameter measurement for each individual tree, the tree crown polygons were used to clip the LiDAR point cloud data such that the points within each polygon clouds were assigned the same ID as the individual tree crown polygon ID. Individual tree LiDAR metrics were then computed using the rMetrics function in the rLiDAR package (Silva et al. 2015). The generated metrics from LiDAR were used to model the individual tree heights represented by the LiDAR height metrics (Figure 4.2).

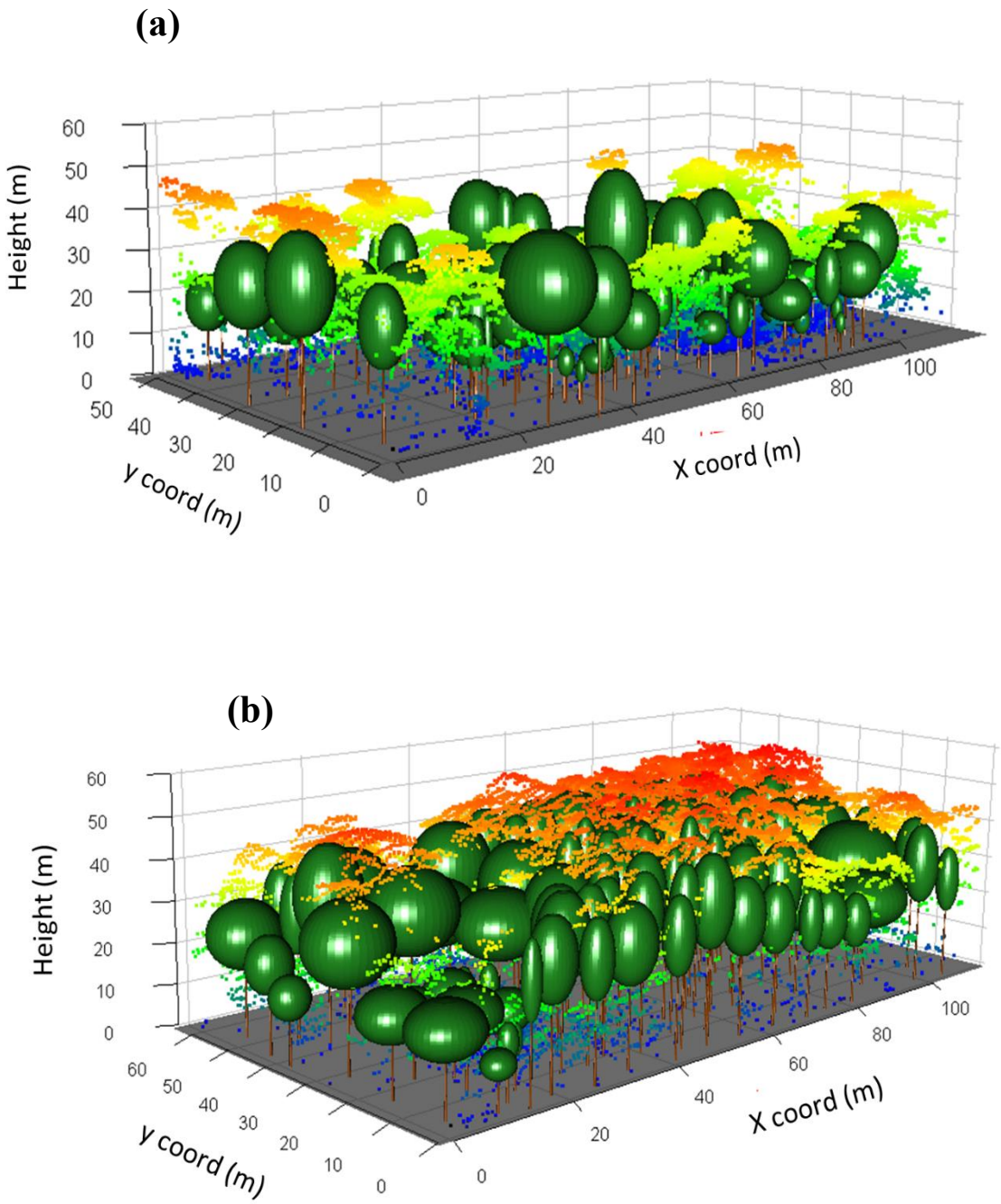


Figure 4.2. Dimensional Tree Height Model. Individual tree crowns extracted from the LiDAR point clouds (overlaid) at (a) PFR and (b) FRIM.

**Note:* Trees measured in PFR site were based on selected criteria as explain in section 4.2. Not all trees were measured as this explained why some trees are represented only by the point cloud. The tree models are ellipsoids based on field data.

4.3.3 LiDAR Metrics

Discrete return LiDAR metrics are descriptive structure statistics that are calculated from the measurement of the height normalized point cloud in three-dimensional space (Lefsky et al. 2005). In this study, 30 metrics for each individual tree were calculated including; (i) selected percentile heights (i.e., 5,10,….99 denoted as h_5,h_{10},….,h_{99}, respectively), maximum height (h_{max}), crown base height (CBH), mean height (h_{mean}), median height (h_{med}), mode height (h_{mode}) and (ii) variability of height measures, i.e., coefficient of variations of height (h_{cv}), variance (h_{var}), kurtosis ($h_{kurtosis}$), skewness ($h_{skewness}$) and standard deviation (h_{sd}). Crown Width (CW) were computed using the chullLiDAR 2D function after computing the canopy area. K- Means cluster was used to remove the points below the crowns as adapted from Dean (2009) approach to calculate CBH metrics. To examine the impact of laser returns on AGB estimation, point cloud LiDAR metrics based on all returns within a tree polygon were generated (Table 4.2).

Table 4.2. List of LiDAR metrics

Metrics	Description
Total number of returns	Total number of discrete LiDAR measurements for individual tree
Crown Base height (CBH)	Points height above the crown base height
Maximum height (hmax)	Maximum height above ground of all LiDAR returns for individual tree
Mean height (hmean)	Mean height above ground of all LiDAR returns for individual tree

Median height (hmed)	Median height above ground of all LiDAR returns for individual tree
Mode Height (hmode)	Mode height above ground of all LiDAR returns for individual tree
Standard deviation (hsd)	Standard Deviation of heights of LiDAR returns for individual tree
Percentile height (h5, h10,h20,h25,h30,h40,h50, H60,h70,h75,h80,h90,h95,h99)	The percentiles of the canopy height distributions (5th,…,99th) of all returns
Hcv	Coefficient of variation of heights of all LiDAR returns
hvar	Distribution of average height variance derived from all LiDAR returns
hkurtosis	Distribution of height kurtosis
hskewness	Distribution of height skewness
CW	Crown Width

4.4 Statistical Analyses

Assumption of homogeneity of regression slopes in analysis of covariance (ANCOVA) was tested before developing the full model. This is to ensure that field sample data from FRIM could be pooled with PFR for regression model development. To estimate the AGB in PFR, the regression model that represents the relationship between the selected LiDAR metrics and the AGB was developed. The statistical analysis involved three steps: (1) selection of independent variables, (2) regression model development, and (3) model assessment.

4.4.1 ANCOVA

Since AGB at FRIM differed from PFR due to the different stand conditions, it is important that data pooling does not compromise the LiDAR-AGB relationship for the

regression model in the study area. This can be tested using an ANCOVA approach (Li *et al.* 2010). The AGB was estimated using the selected appropriate allometric equations and field data. Initially a linear regression model was chosen to observe any non-random patterns in the residuals. A curve was fitted to the trends of residuals for both forests.

ANCOVA was conducted with a general regression model having one continuous outcome variable and one or more factor variables. It tests whether certain factors have an effect on the outcome variable after adjusting for the effects of confounding factors. In this study, ANCOVA was used to determine whether the sampled data from PFR and FRIM forest could be pooled for the AGB regression model in representing both study areas through testing the assumption of homogeneous regression slopes, which assumed that the relationship between the AGB and the LiDAR metrics in PFR and FRIM were independent of the regional conditions. To test this assumption, interaction effects were added into the regression model by adding the product of covariates and the regional factor. A power transformation was used to test this assumption because it best fits the sample data in the study:

$$Y = (\beta_0 + \beta_{site}) + (\beta_1 + \beta_{1site}) \text{ Ln CW} + (\beta_2 + \beta_{2site}) \text{ Ln (h80)} . \quad (2)$$

Where Y = field values of AGB (Kg/tree); *site* is a dummy variable for representing the different site (1= FRIM, 0= PASOH) and CW and h80 is the independent variables. CW and h80 was used to test this effect because this is the best combination variables to build up AGB model as explained in detail at Table 4, section 4.5.1. When site is zero, the intercept and the slopes become $\beta_{0,}$ β_1 and β_2 respectively in the model, which

is derived only from PFR. When site is 1, the intercept and the slopes of the model include the addition of field sample data from FRIM. This approach was used to determine whether adding the sampled data from FRIM would impact the intercept and the slopes of the model significantly. The null hypothesis for the F test that $\beta_0 = 0$, $\beta_1 = 0$ and $\beta_2 = 0$; adding the FRIM data changed neither the intercept nor the slopes.

4.4.2 Selection of independent variables

The strength of relationships between AGB and the 30 LiDAR metrics were tested with the coefficient of determination (R^2). Regression models with highly correlated independent variables are not stable from a statistical perspective and are hard to interpret from a biological perspective (Li et al. 2008). Following the variables selection method of Næsset et al. (2005) who minimized the number of LiDAR metrics to avoid information redundancy and promote parsimony, the original LiDAR metrics were reduced to non-correlated principle components. It is difficult to interpret the principle components themselves because they are in linear combinations of the original LiDAR metrics and do not themselves have a clear physical meaning (Li et al. 2008). Variance Inflation Factor (VIF) estimates how much variance of a regression coefficient is inflated due to multicollinearity in the model. Some of the metrics selected introduced serious multicollinearity with VIF value greater than 3. The solution was to shift to the variable selection method of (Næsset & Gobakken, 2008) where all of the independent variables were included as possible predictor variables for selection using both stepwise variable selection with an R^2 improvement technique.

The model simulates the relationship between field AGB and LiDAR variables in all sites and in all combinations. The estimated AGB, a dependent variable, was calculated based on the converted ground measurements taken during October 2014 at PFR and April 2014 at FRIM (see Chapter 3). All 30 variables of LiDAR metrics were used as potential model-independent variables.

4.4.3 Regression model development

Power functions have been used in a large number of studies for AGB estimation, probably because many allometric equations for calculating AGB in the field are power functions, in addition to some theoretical explanations of why allometry works at all (e.g. Niklas, or West, Brown, Enquist). The relationship between AGB and forest height has been well described (Morel et al. 2011; Wang et al. 2013), and the power function is widely accepted (Niklas & Enquist, 2001) as:

$$\mathrm{AGB} = a\,(\mathrm{H})^{b}\,. \qquad (3)$$

where AGB is the plot AGB (Mg/ha), H the field tree height (m) and a and b are coefficients

The allometric equations used to derive AGB from field data in this study were based on a power function developed by (Chave et al. 2014) and thus, provide a solid justification for the chosen model. Because most allometric equations for calculating tree-level AGB from field measurements are power models, AGB and LiDAR metrics were log transformed when fitting a regression model. This is to reduce the heterogeneity of the regression residual variance. A multiplicative model formulated

as eq. (3) was selected, which can be translated into a linear form according to eq. (4). This type of model has been used successfully by others to model various forest biophysical properties (Naesset 2002; Lim et al. 2003). The natural logarithm transformations required in Equation (4) also ensure that, in most cases, regression assumptions are not violated. In the regression analysis, a linear multiplicative model was used which correspond to multiplicative power transformation, as used successfully by others to estimate various forest attributes. All derived variables were transformed to the natural logarithm. First, AGB was regressed against height metrics and AGB as a dependent variable. Multiple linear regression analysis using all independent variables was then carried out for both sites independently and in combination. Stepwise selection using Akaike Information Criterion (AIC) (Akaike 1974) was performed in R to select variables included in the final model. AIC is a goodness of fit measure that favors smaller residual error in the model, but penalizes for including further predictors and helps avoid overfitting. At each step, individual variables were either added or deleted and the next model with lowest AIC was retained in the next step. The coefficient of determination (R^2), the adjusted coefficient of determination (Adj-R^2) and the root-mean-square error (RMSE) were calculated for model comparisons.

$$Y = \beta 0 + \beta_1 CW + \beta_2 hmax + \beta_3 CBH + \beta_4 h_5 + \ldots\ldots + \beta_{30} h_{99} \quad . \qquad (4)$$

$$\ln Y = \beta_0 + \beta_1 \ln CW + \beta_2 \ln hmax + \beta_3 \ln CBH + \beta_4 \ln h_5 + \ldots.. + \beta_{30} \ln h_{99} \quad . \qquad (5)$$

Where Y = field values of AGB (Mg ha^{-1}); hmax = maximum height of canopy; CBH = Crown Base Height, h5,…, h99 = percentiles corresponding to 5,…, 99% of the laser canopy height (m); CW = crown size width. Both stepwise variable selection and the

maximum R^2 variable selection techniques were applied to select LiDAR variables to be included in the models. When using the stepwise selection method, no independent variables were left in the models with a partial F statistic significance level greater than 0.05. The best fitting models were selected based on the lowest AIC value. The variance inflation factor (VIF) was used to address multicollinearity issues by calculating and monitoring the size of the condition number. Because the best models suggested by the stepwise procedure might have many independent variables that could introduce multicollinearity problems, the maximum R^2 improvement technique searches for the "best" one-variable model.

In accordance with the objectives of this study, the influence of the forest site on the estimated AGB models were assessed by extending the preliminary regression models derived above with dummy variables representing these factors. Since both sites are lowland dipterocarp forest with almost similar species, the homogeneity test to assess whether both datasets can be pooled for model development was only assessed on variables related to site properties. To assess the effects of different forest site, the dummy variables were assigned a value of 0 for sampled data in Pasoh and a value of 1 for sampled data in FRIM. This result will be further discussed in ANCOVA analysis in ANCOVA.

4.4.3.1 Model assessment

Two questions were essential in assessing the model: (a) how well does the model fit the sampled data (model fitting analysis), and (b) is the model generalizable outside

the sampled data? In the model fitting analysis, potentially influential data outliers were identified using the studentised deleted residual and Cook's distance statistic (Li *et al.* 2010) to assess a data point's influence on the regression coefficients of the regression model; generally, it should be considered a potentially influential point when its value exceeds 1. The goodness of fit of the model was evaluated using an adj-R^2 and the RMSE, and standardized residuals were used to check the following model assumptions: (1) the equal variance for all independent data, and (2) no systematic pattern between the regression model residuals and the predictions. R^2 is the statistic for evaluating the fit of a linear regression model to the sampled data. However, the impact of the degree of freedom on the model accuracy is well known: the more predictive variables in the model, the higher the R^2. Clearly, the R^2 is not a good index to assess the model's goodness of fit. An Adj-R^2 removes the impact of the degree of freedom, and thus provides a more conservative measure of the model's goodness of fit.

The assumption of normality of error terms were assessed using the Shapiro-Wilk test (Shapiro & Wilk 1965; Kramer et al. 2016; Silva et al. 2016) and heteroscedasticity were tested using the Breusch-Pagan test (Breusch & Pagan 1979; Kramer et al. 2016; Silva et al. 2016). Comparisons between models were based on their predictive capabilities with respect to the coefficient of determination (R^2), root mean square error (RMSE) and AIC between observed and predicted values. Once the best model was chosen, 10-fold cross validation was performed. RMSE were used as a way to measure values measured and the true value being estimated. RMSE was calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad . \qquad (6)$$

Where y_i is the observed values and $\hat{y}_i$ is predicted values for the *i*th compound, respectively, and *n* is the number of samples in the training set.

4.5 Results

A linear regression model was chosen to evaluate whether a relationship exists between AGB and LiDAR variables. Residuals vs. modeled AGB were plotted to examine the assumption of linearity. To help observe patterns of residuals, a curve was fitted to the trends of residuals (Figure 4.3). The plot clearly displayed non-linear trends in the residuals.

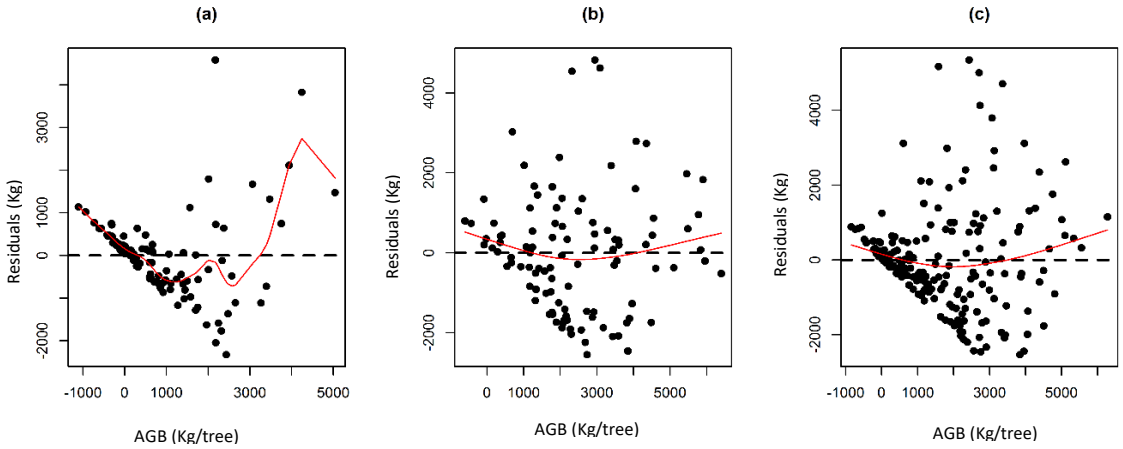


Figure 4.3. Residuals on fitted linear regression; (a) PFR site; (b) FFR site and; (c) pooled data between PFR and FFR site – all indicating randomness existed in their relationship.

4.5.1 Independent variables selection

Regression models with log-transformed variables were estimated with LiDAR derived metrics as the only independent variables. The models were selected by a

stepwise regression procedure using all possible independent variables from both sites independently and combined. These are summarized in Table 4.3. To simplify the models, the maximum R^2 improvement technique was used to find the "best" variable combination and the "best-two" variable model and so- forth, with the benchmark of the best fitting model being the one with the lowest AIC value. Multicollinearity was further evaluated to confirm that all independent variables had correlations below 0.90, a VIF value below 5, and high correlation with AGB.

Table 4.3. Top four best independent variable combinations with AGB as the dependent, based on lowest AIC values, from two forest sites considered independently and in combination.

Study Site and Model	**Selected Independent Variables (Natural Log transformed)**	**Adj-R^2**	**AIC**	**RMSE (Kg/tree)**	**rRMSE**
PFR	1) CW, CBH, h70	0.74	256.45	0.79	13.65%
	2) CW, h70	0.74	256.72	0.79	13.74%
	3) CW, CBH, h5, h70	0.74	257.05	0.78	13.61%
	4) h90	0.61	298.71	0.97	16.84%
FRIM	1) CBH, hmode	0.25	300.10	1.00	13.73%
	2) CW, hvar, hcv, h90	0.22	305.85	1.01	13.96%
	3) h90	0.11	316.81	1.09	14.96%
	4) hmedian, h45, h55	0.12	317.70	1.08	14.86%
Pooled Data site	1) **CW, CBH, h5, h75**	**0.63**	**579.65**	**0.95**	**14.52%**
	2) **CW, h80**	**0.63**	**581.90**	**0.96**	**14.68%**
	3) CW, CBH, h5	0.63	591.15	0.97	14.96%
	4) h90	0.53	627.68	1.12	16.41%

Based on high Adj-R^2 and low rRMSE values, the predictive models combining data from two forest sites proved efficient strategy to predict AGB on the region. The best model was selected based on lowest AIC value, however, the selection of the best variables was determined by the correlation coefficients and p-values of the partial F statistic of selected metrics. Figure 4.4 shows a scatterplot matrix of the two models

with lowest AIC value. The model combining four variables (CW, CBH, h5 and h75) gives the lowest AIC value and highest Adj-R^2, but this combination introduces multicollinearity with VIF value > 5 and high inter-correlated independent variables with r > 0.90. This value can be summarized through Figure 4.4(a). The model combining the CW and h80 variables, however, passes all the diagnostic tests with VIF < 3 and low correlation between predictors with r < 0.70 used for further analysis in this study. This can be summarized through Figure 4.4(b). The histogram of each metric is drawn in the diagonal line. The kernel density overlaid and the significant asterisks shows the level of significant (*0.05, **0.01 and ***0.001). Further model assessment will focus on the model based on the combined sites instead of the individual test sites.

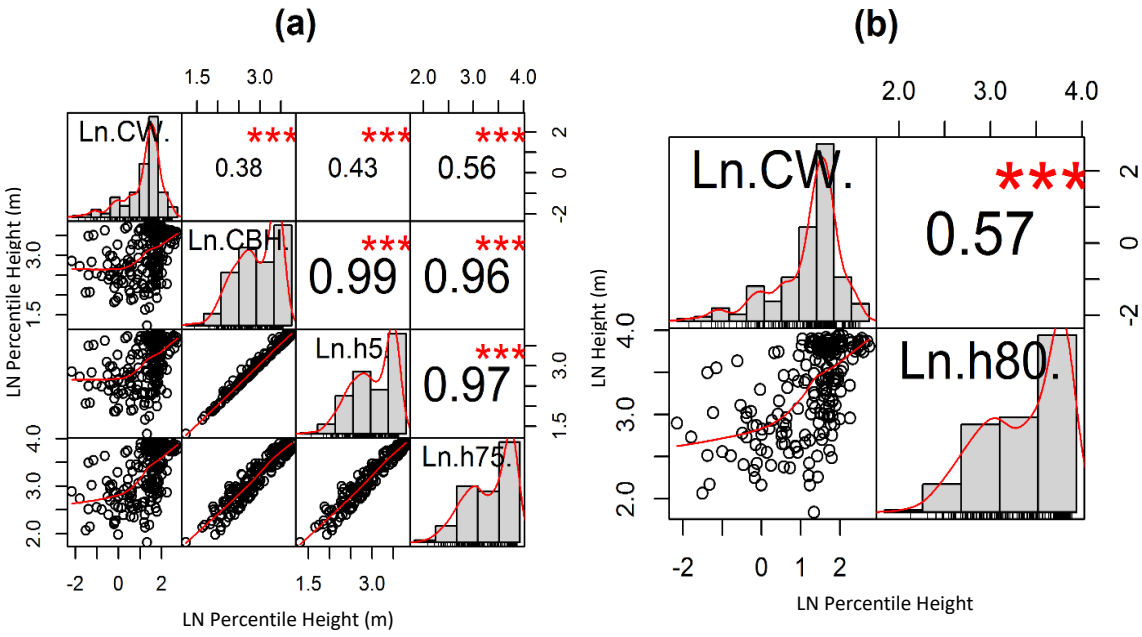


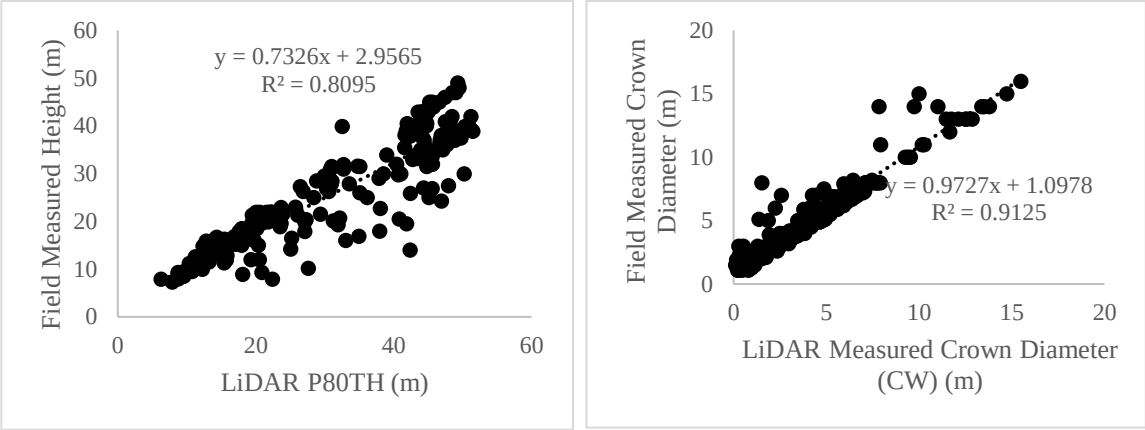
Figure 4.4. The matrix of scatter plots (lower panel) and the correlation coefficients (r) (upper panel) for all detected trees from all possible best metrics in pooled data site. (a) High correlated metrics (with r>0.90) for model with CW, CBH, h5 and h75 as the independent variables and (b) Low-correlated metrics (with r<0.7) for model with CW

and h80 as the independent variables for all detected trees. The histogram of each metric are drawn in the diagonal line. The kernel density overlaid and the significant asterisks shows the level of significant (*0.05, **0.01 and ***0.001)

From the selected independent variables, an empirical approach was employed to identify the most appropriate curve that fitted the data. Linear, Cubic, Quadratic and Power curve functions were fitted to the data. The R^2, Adj-R^2 and RMSE were used to determine the most appropriate model. Of the four non-linear functions tested, the power function was found to best fit the sample data (Table 4.5). The field measured tree height and h80 of 209 total trees had an R^2 of 0.81 and a strong correlation (R^2 = 0.91) was also obtained from the relationship between field measured crown diameter and the diameter measured on the trees identified with LiDAR (Figure 4.5).

Table 4.5. Comparison of the curve fitted transformation in nonlinear regression model from pooled sample data between independent variable and dependent variable (AGB).

Independent Variable	Model	R^2	Adj-R^2	RMSE (Kg/tree)
CW	Linear	0.385	0.382	1745.172
	Cubic	0.395	0.386	1739.572
	Quadratic	0.386	0.380	1748.366
	Power	0.450	0.447	1.161
h80	Linear	0.245	0.241	1933.415
	Cubic	0.254	0.244	1930.686
	Quadratic	0.251	0.244	1929.913
	Power	0.532	0.530	1.071



(a) (b)

Figure 4.5. Relationship of field measured height (m) with LiDAR h80 (a) and (b) relationship of field measured crown diameter (m) with LiDAR measured crown diameter (m)

4.5.2 ANCOVA for pooling data from different forest sites

The regression output of Eq (2) fit to the 209 individual trees from both sites and the F test for the overall model was statistically highly significant ($p \leq 0.001$) and provided a good fit to the data (Table 4.6). It was found that all the interaction effects, Ln (CW) * dummy site and Ln (h80) * dummy site were not statistically significant according to their p values (0.782 and 0.339 respectively), which supported the null hypothesis: the response variables of the AGB in PFR and FRIM were independent of the site factor. Therefore, there were not statistically significant changes in the regression slopes of the model in the presence of different data from FRIM site. This result confirmed an early assumption that the vegetation structure was likely to have more impact on the LiDAR metrics than site-level environmental factors. Therefore, the data from FRIM could be pooled with the PFR data to develop the regression models and to mitigate the issue of the limited sample size in this study.

Table 4.6. Result of analysis of covariance derived from 209 individual trees

Source	Sum of squared (Type III)	Df	F value	Sig.
Intercept	0.044	1	0.049	0.8250
Ln (CW)	39.546	1	43.861	3.04e-10
Ln (h80)	71.087	1	78.844	3.41e-16
Ln (CW) * dummy site	0.069	1	0.077	0.782
Ln (h80) * dummy site	0.827	1	0.917	0.339

*Residual standard error = 0.9495 on 205 degrees of freedom, Adj-R^2 = 0.630, F-statistic: 90.04 on 4 and 205 DF, p-value: < 2.2e-16

4.5.3 Data outlier and influential analysis

The results from the regression are shown in Table 4.7. The p-value and R^2 value for the model were ≤ 0.001 and 0.61, respectively; thus, the model, provided a good fit to the data.

Table 4.7. Results of nonlinear power model

	Coefficients	Estimated Std. error	t value	Pr(>|t|)
(Intercept)	0.612	0.463	1.324	0.187
Ln (CW)	0.629	0.085	7.375	3.9e-12
Ln (h80)	1.540	0.154	10.012	<2e-16

*Adj-R^2 = 0.626, residual standard error = 0.9553 on 207 degrees of freedom, F-statistic = 175.7 on 2 and 207 DF, p-value: < 2.2 e-16

By examining the influence of outliers and diagnostic plot after fitting the regression model, there were clearly no pattern of violations and the fitted model works well with the data (Figure 4.6). The diagnostic works by first examining the residuals vs fitted values (Figure 4.6a). The smoothed curved red line that passes through the actual residuals and lies close to the gray horizontal dashed line with a relatively flat pattern (residuals range ± 0), indicating a good fit to the data. Several points numbered in the

plot are the points that should be examined because not all points necessarily indicate a problem. Tree ID 112, 185 and 200 are taken for further examination.

Normal probability plot of the residuals evaluates the assumption that errors are normally distributed through where the points lies very close to the dashed line (Figure 4.6b). Some deviation occurs particularly near the start and the ends. The x-axis in Figure 4.6(c) is identical to the x-axis on Figure 4.6(a), and the y-axis is the square root of the standardized residuals, which are residuals rescaled so that they have a mean of zero and variance of one, and all values are positive. This plot eliminates the sign on the residual, with large residuals (both positive and negative) plotting at the top and small residuals plotting at the bottom. The red line shows the trend. The regression assumed homoscedasticity such that the variance in the residuals doesn't change as a function of x. The assumption is correct except towards the end at the far right end, where a few data points pull it down. This is due to the residuals being too spread along the x-axis as it passes around 4. Because the residuals spread wider and became clustered at the end, the red smooth line is not horizontal and shows a steep angle down.

The influence of each observation on the regression coefficients were then examined through Cook's distance plot (Figure 4.6d). The Cook's distance statistic is a measure, for each observation in turn, of the extent of change in model estimates when that particular observation is omitted. Cook's distance is always positive but with no upper limit. Läuter (1985), suggested that Cook's distance values greater than 1, or that are

substantially larger should be examined. From this plot, tree IDs 54, 112 and 126 are more influential than other trees and should be investigated further in terms of residual and hat values. Standardized residuals against leverage (Figure 4.6e), shows that standardized residuals are centered and symmetrically around zero, expected for a normal distribution. Leverage is a measure of how much each data point influences the regression. Because the regression must pass through the centroid, points that lie far from the centroid have greater leverage, and their leverage increases if there are fewer points nearby. As a result, leverage reflects both the distance from the centroid and the isolation point. The plot also contours values of cook's distance, which measures how much the regression would change if a point was deleted. Cook's distance is increased by leverage and by large residuals: a point far from the centroid with a large residual can severely distort the regression. On this plot, the red smoothed line stays close to horizontal gray dashed line and that no points have a large cook's distance (>0.5). This relationship can be summarized in plot 4.6(f).

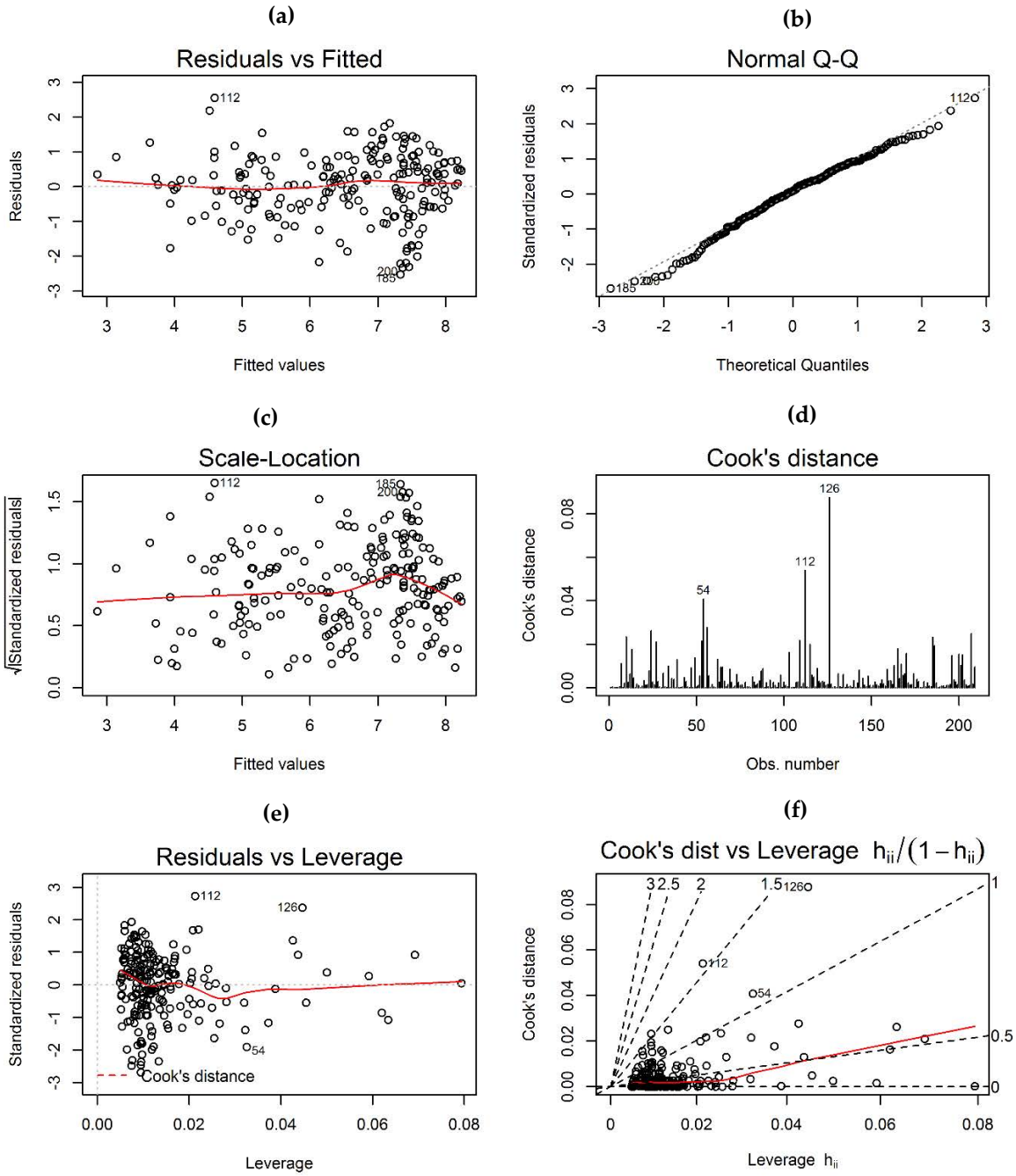


Figure 4.6. Diagnostic Influence plot of residual for the pooled data site. (a) Residual vs fitted plot, (b) Normal Q-Q plot, (c) Standardized residual plot, (d) Cook's Distance plot, (e) Residual vs Leverage plot and (f) Cook's Distance vs Leverage summary plot.

R has its own rules for flagging points as being influential because R produces a set of standard plots for linear regression model that help to assess whether the assumption

are reasonable or not. Studentised residual, Hat as a measure of leverage and cook's distance were used to examine the outlier (Table 4.8).

Table 4.8. Influence measure for AGB model

Tree ID	Site	Height (m)	DBH (cm)	Wood density (g/cm^{-3})	AGB values ($MgHa^{-1}$)	StudRes	Hat	CookD
54	Pasoh	13.90	11.60	0.57	8.702	-1.908	0.033	0.041
112	**FRIM**	**15.60**	**40.30**	**0.96**	**1260.451**	**2.727**	**0.021**	**0.054**
126	**FRIM**	**12.00**	**27.50**	**0.96**	**814.283**	**2.372**	**0.045**	**0.088**
185	FRIM	27.00	14.00	0.53	122.078	-2.693	0.010	0.023
200	FRIM	20.50	11.10	0.57	152.440	-2.486	0.007	0.016

ID 126 represent individual tree number 1106184 from FRIM forest has fairly moderate leverage, a relatively low residual and moderately high influence. ID 112 represent individual tree number 1106124 from FRIM Forest, has small leverage and relatively small residual but very little influence. ID 54 which represent tree number 62 from Pasoh forest, ID 185 represent tree number 1107090 and ID 200 represent tree number 1107195 both from FRIM forest were excluded from this diagnostic as the influence is relatively low. Even though point ID 126 shows highest coefficient in all influential test, diagnosed in this study was considerably too small to break the assumption and will not give a significant influence on the model. Therefore, it was reasonable to retain this point in the model.

4.5.4 Regression Model fitting analysis

Multicollinearity is a serious issue that must be considered when using regression models. Multicollinearity was controlled by checking the VIF of the models. All independent variables and the model as a whole had a VIF below 5. Normal probability plots of the residuals suggested that the residuals were normally distributed. Besides the normal distribution Q-Q plot, residuals from the prospective model were also tested for normality by using the Shapiro-Wilk test and analysis on heteroscedasticity by

using the Breusch-Pagan test. Both tests accept the null hypothesis that residuals from the model are normally distributed by reporting p-values = 0.13 for Shapiro-Wilk test and the variance of the error term is constant (homoscedasticity) with p = 0.34. To further examine the regression model after transformation, Figure 4.7 shows the residuals vs. fitted values (predicted). The residual plots show approximated mean and spread of points at each fitted value and the error centered around zero. The plot look symmetrically distributed along the y-axis.

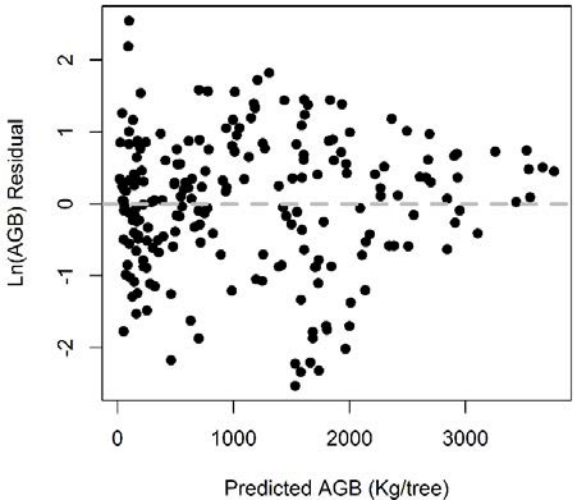


Figure 4.7. Standardized residual plot of the final model from combined data set.

By adopting the power transformation into the model's design, the residual variances were reduced and the nonlinearity was addressed. Further, the linearity was not violated according to the normal probability plot of the residuals and the power function model was best fit to the sample data. Next, how well the proposed model could predict the outcome in different datasets (known as model generalizability) need to be assessed. The generalization of the proposed regression model was evaluated using the 95% confidence intervals of the prediction in addition to cross-validation. The model performed well for small predicted values (Figure 4.9).

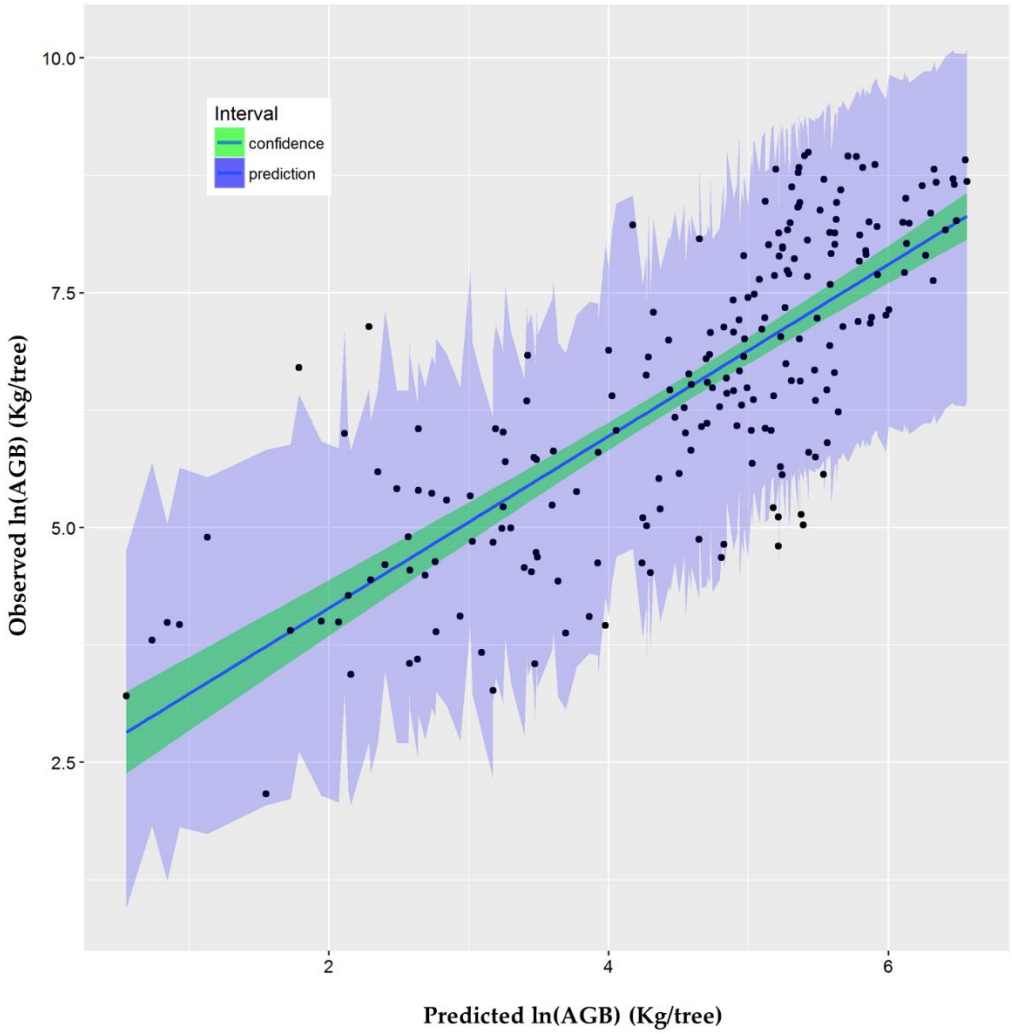


Figure 4.8. 95% confidence intervals of the predictions of power regression model. 95% probability that the true best-fit line for AGB lies within the confidence interval and 95% of the Y-values to be found for a certain X-value within the interval range around the linear regression line.

The mean square error (MSE) from a ten-fold cross validation was 0.907 compared to 0.856 that was estimated from the 105 samples (Table 4.9). Since the two results were reasonably close, the proposed model can be considered generalizable. The MSE from prediction model were substituted into the equation to calculate correction factors of 1.46; after multiplied with the back-transformed predictions, these correction factors slightly underestimated the mean AGB by 0.69 kg/tree (Table 4.10). In general, the

distribution of predictions better matched the distribution of observation after bias correction than before bias correction. Figure 4.9, shows the model's predictive results for both the training data and calibration data sets in each pass.

Table 4.9. Assessment of the predictive accuracy of the regression model by 10 fold cross-validation

Observations	Sum of Squares	Mean Square	Number of sample (n)
Fold 1	13.90	0.70	20
Fold 2	20.30	0.97	21
Fold 3	17.10	0.82	21
Fold 4	17.10	0.82	21
Fold 5	12.80	0.61	21
Fold 6	13.10	0.62	21
Fold 7	21.30	1.02	21
Fold 8	29.70	1.41	21
Fold 9	22.40	1.06	21
Fold 10	21.80	1.04	21
Overall sum over all 21 folds		**0.907**	

Table 4.10. Summary statistics of observed and predicted AGB (Kg/tree) before and after bias correction for the inverse natural logarithm transformation

AGB	Min	First quartile	Median	Mean	Third quartile	Max
Observations	8.70	199.29	706.05	1666.69	2684.35	8067.16
Predictions after back-transformation	17.45	248.10	934.70	1138.00	1735.00	3763.00
Predictions corrected for bias transformation	25.54	363.10	1368.00	1666.00	2539.00	5508.00

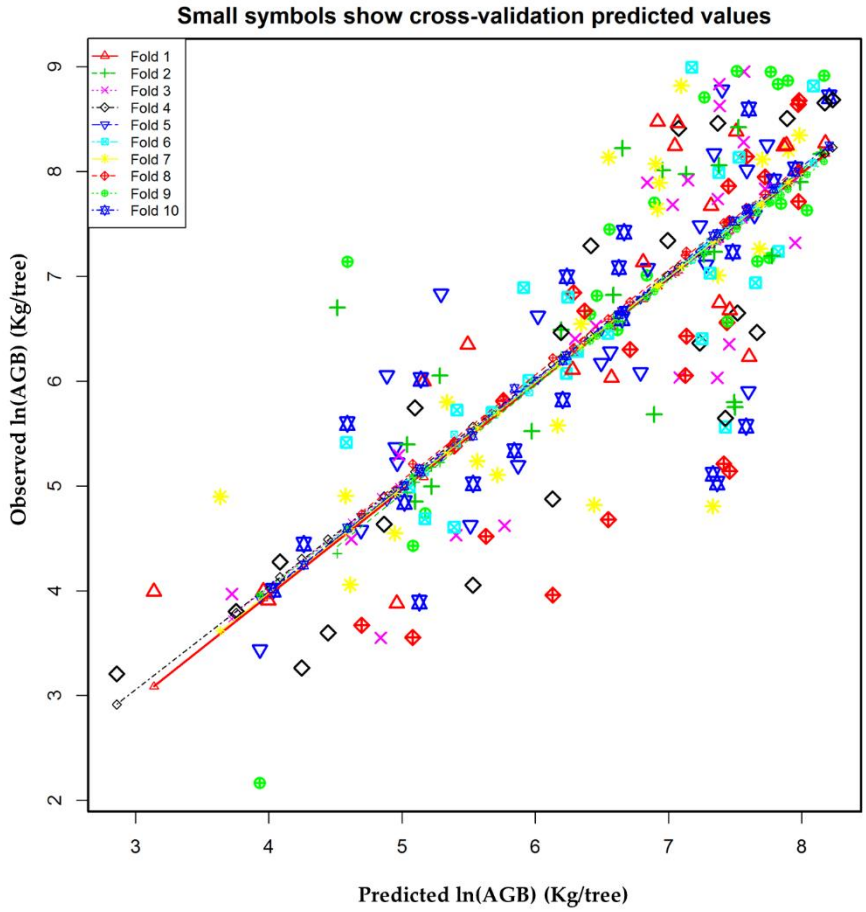


Figure 4.9. Summary of 10-fold cross validation for evaluating AGB power model. The line is fitted to the training data set in each pass, leaving out corresponding test data set. Prediction of the omitted test data are used to assess the predictive accuracy. The dashed lines are parallel and close to each other. The model's curve was scattered gradually in the low AGB values and close to each other towards the high AGB range, slightly show a positive trend of errors in the low AGB range.

AGB from the two sites was well estimated using the final prediction model predicting AGB from two LiDAR metrics (i.e. CW and h80), with Adj-R^2 of 0.63 and RMSE of 14.68% (Figure 4.10).

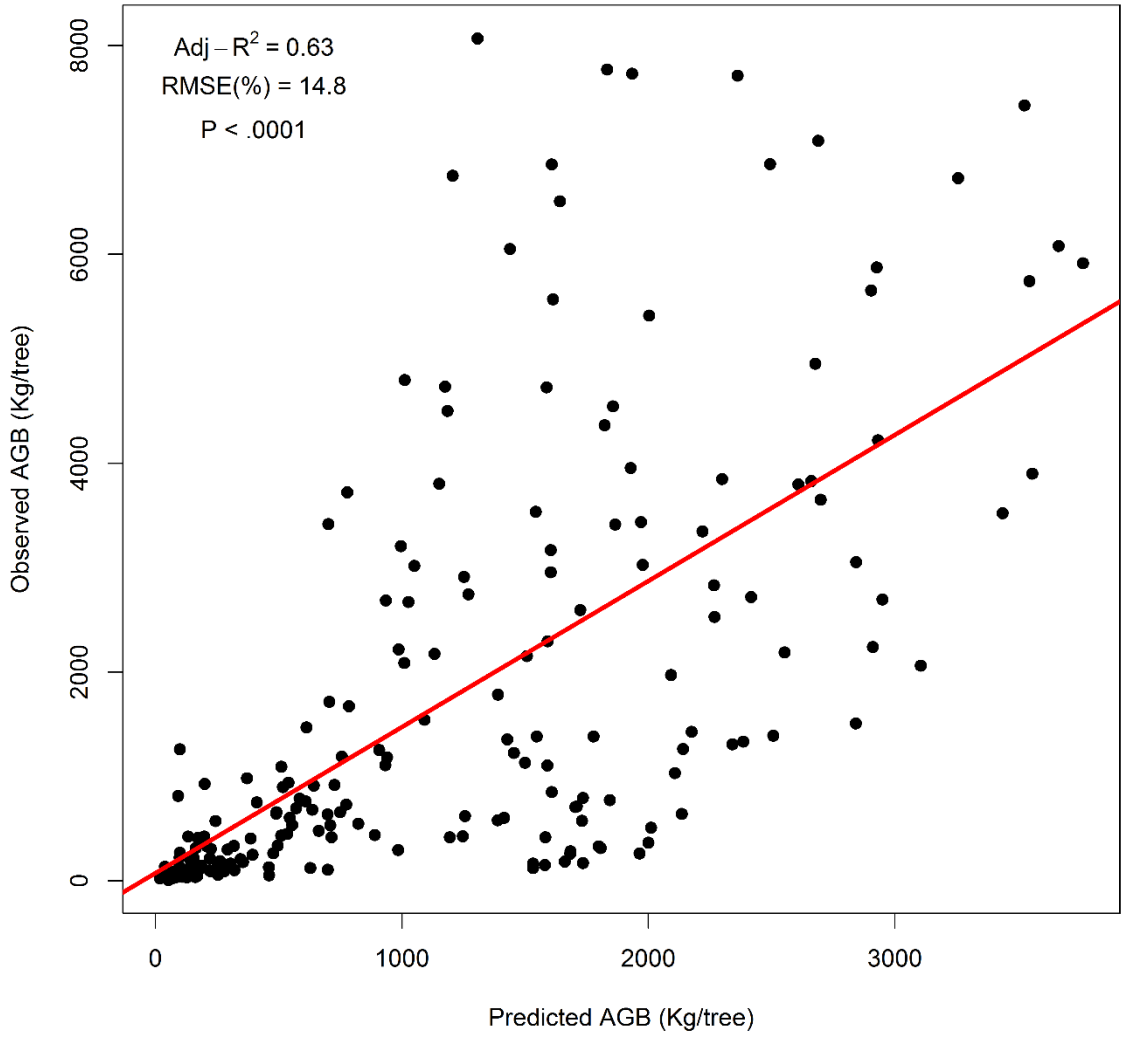


Figure 4.10. Scatter plot of final observed *versus* predicted AGB from the combination at both sites

4.6 Discussion

Malaysia's interest in participating in the World Bank's Forest Carbon Partnership Facility (FCPF) (Omar 2012) requires that they have a baseline study to calculate carbon so as to present country's current status of carbon stored in forest. The development of biomass model for carbon stock estimation will help to support the research development for carbon monitoring methodology in Malaysia with aims to

prepare forested countries for REDD+ implementation in Malaysia (FRIM 2011, FRIM 2012). To the best of the author's knowledge, this is one of the first studies to develop a model using individual tree and LiDAR-derived crown metrics at tree level in tropical rainforest. The increasing importance of accurate biomass estimation to support the REDD+ implementation, has created a critical need to understand, evaluate and improve current tree biomass prediction methods by adopting state-of-the-art analytical and statistical techniques. In this study, LiDAR metrics that correlate well with field AGB were extracted based on trees that were mapped in the field as the main point to develop the LiDAR-AGB model. It should be stressed that the method used to extract LiDAR data for the trees was not an automatic individual tree detection method, but rather it was based on trees that were mapped in the field. Chapter 5 discussed the automatic individual tree detection approach in detail.

4.6.1 Field-LiDAR AGB model

Field sample data can be valuable for evaluating LiDAR-based and other remotely sensed AGB maps, as plots are systematically arranged to provide a spatially unbiased estimate of forest AGB over an area, followed by well-documented measurement protocols that are quality controlled (Johnson et al. 2014). Findings from this study show that the LiDAR metrics and stem-localized AGB correlated very well with <15% error in the residuals. An interesting finding from this research was that crown width (CW) was one of the best LiDAR metrics for predicting AGB. Incorporating both crown size and tree height, particularly of large trees, may improve estimates from remote sensing data for both standing carbon stocks and carbon stock changes and this may be especially applicable to methods based on small footprint LiDAR (Goodman

et al. 2014). A recent published study by Ferraz et al. (2016) also incorporates CW, from decomposed entire point cloud into 3-Dimensional (3D) clusters that correspond to individual tree crown. The 3D clusters were modelled using a convex hull to calculate the crown area (CA) and the crown volume (CV) using the same tools developed by Silva et al. (2015) as used in this study. However, the method allows for the use of many forest parameters in existing field allometrics equations to estimate the AGB. This method is only appropriate for forest areas with well-established field AGB allometry. The model developed based on a geolocated stem map that represents most of the species which exist for lowland dipterocarp forest for building up LiDAR-AGB model with an adequate comparison of individual tree at stem level was presented. Developing a regression model for AGB estimation at stem level is important to derive the most important LiDAR metrics that correlate with field AGB before this could be implemented to landscape level. The initial approach for quality assessment was presented in this study.

CW is related to DBH (Condit 1995; Gering & May 1995; Sumida et al. 2013), direct retrieval of DBH from LiDAR point cloud at tree level is possible at temperate forest, but this has not yet been addressed in a substantial way (Maltamo et al. 2009, Vauhkonen et al. 2010). However, it seems it is not possible to apply the approach to tropical rainforest. Instead, an interpolation through the relationship with the crown size is needed. Given that extraction of CW can be achieved through LiDAR, it is possible that future studies could retrieve DBH using similar methods. The CW and DBH relationships of tropical forest has previously been reported in (Kwan 1966, Perez 1970), the CW and DBH relationships has also been reported by Wile (1964),

Roberts & Ross (1965), and Bonner 1968) for a number of conifer species. All of these studies found a strong relationship between DBH and CW (R^2= 0.6 – 0.9).

4.6.2 Metrics selection and their explanation

The analysis involved all the independent variables in the initial search - finding correlations among independent variables, starting the analysis using the less correlated metrics (r<0.7) and discarding those with least importance until the classification accuracy became stable. The top-ranked metrics were selected based on the lowest AIC value as the first criteria. To make sure the best model with the lowest AIC value is stable from the statistical perspective, multicollinearity was tested using VIF. The best model based on AIC contained CW, CBH, 5th percentile and 75th percentile. However, it also introduced high multicollinearity based on VIF value >5 and the p-values are not significant for CBH (0.06) and 5th percentile (0.32) variables. There were also strong high correlations among independent variables as shown in Figure 4.4(a). Regression models with highly correlated independent variables are not stable and hard to interpret from either statistical or biological perspectives (Naesset & Gobakken 2008). The second best model with the lowest AIC value was the model containing CW and 80th Percentile height as the independent variables for predicting AGB. This model passed all the diagnostic tests, with all variables having a VIF less than 5 suggesting no serious multicollinearity in the model. In addition, all variables were significant in terms of p-values, and there were no statistically significant changes in the regression slopes of the model in the presence of combining data from FRIM field sample plot.

There have been other studies reporting different height metrics that correlate very well with AGB that varies not just depending on forest type and location, but also on model and data processing procedures (Thapa et al. 2015). In this study, multiplicative models using power functions performed best for predicting AGB from metrics generated from the raw point cloud LiDAR. AGB is usually non-linearly related to remote sensing variables; therefore, nonlinear transformations such as a power function where the response and explanatory variables were log transformed (Hall et al. 2005) reduces the heterogeneity of the regression residual variance. On the plus side, power functions have been used in a large number of studies for AGB estimation, probably because most allometric equations for calculating AGB in the field are power models (Chen 2013). An empirical approach has been implemented in this study to identify the most appropriate curve that fitted the data and power model was found to be the best fit for the sample data, providing justification for the model specification.

4.7 Analysis of covariance and regression model

Although environmental and successional differences existed between PFR and FRIM, the regression output of ANCOVA demonstrated that the relationship between AGB and LiDAR metrics was independent of these two sites by confirming the assumption of homogeneous slopes. High AGB is related to the size and structure of trees. The smaller the sample size, the greater the errors will be; limited sample size was overcome by combining the sample data between two sites similar in nature. The 105 samples from PFR were probably too small to represent the tree species and structural diversity that exist in tropical rainforest. To overcome the issue of uncertainty in estimating the regression coefficients and intercepts, pooling sample data from similar

forest types is a means to overcome uncertainty in AGB estimation. The final selected model did not violate assumptions of equal variance and normality.

The approach in this study provides direct retrieval of individual metrics (e.g. tree height and crown size) that allow testing of the hypothesis that LiDAR-based AGB models can be replace or complement ground-based AGB models. Ferraz et al. (2016) implemented a similar approach through direct retrieval of LiDAR individual tree metrics, but the model assessment was dependent on a ground-based AGB model, which is an approach only appropriate for a forest area with well-established field AGB allometry. In contrast, our approach used a precisely geolocated stem map that represents for the majority of tropical rainforest species to develop the AGB-LiDAR model. It is much more reliable as it showed that important LiDAR metrics correlated very well with field-based AGB measurements, before implementing at a landscape level.

4.8 Conclusions

ANCOVA analysis confirmed that the relationship between AGB and LiDAR were independent of the forest site, supporting an early assumption that the characteristics of the vegetation structure were likely to have more dominant influence on LiDAR metrics than environmental and other factors between the study areas. Combining data from two forest site has two benefits; first, tropical forest is a very complicated ecological system with diverse species and different biophysical structures, therefore the 105 tree sampled at PFR were possibly too small to build a reliable regression model to represent the whole population that exist in the region. Since PFR and FRIM

forest is similar in terms of topographic conditions, vegetation cover and land use history, it is useful to combine the datasets for model calibration and validation. Second, in terms of statistics, combining data is a great way to understand the result of applying multiple models and approaches. The regression model in this study indicated a good correlation between LiDAR predictors and AGB, as in many prior studies.

A Power function was identified as the best solution to fit the modelled data. Of 30 LiDAR metrics, Height percentile 80th (h80) and crown width, CW was identified as the best combination of independent variables to predict AGB. An interesting finding from this study, supporting earlier findings of Goodman et al. (2014), is the importance of incorporating crown size and height to improve estimates of AGB carbon in tropical forest, especially as this is based on small footprint LiDAR. This research provides an analytic framework for developing a predictive AGB model from LiDAR and field plot data and will be of value especially for forest resource managers for estimating the AGB in lowland dipterocarp forest to improve management decisions. The results of this study, may have been affected by the manual tree delineation, and individual tree position on the field recorded by handheld GPS, however this approach risks giving an over-promising impression of the methods. It is suggested that an automatic procedure with focus to derived multi-layered crown delineation will be a promising avenue for future research. The usefulness of producing accurate tree-level data by means of LiDAR should therefore be assessed carefully with respect to alternative methods, model improvement and the costs involved.

CHAPTER 5

Framework for improving the accuracies of Individual Tree Crown Delineation using Airborne LiDAR data

Wan-Mohd-Jaafar W.S., Woodhouse I.H., Omar H., Silva C.A., (2017). Framework for improving the accuracies of individual tree crown delineation using Airborne LiDAR data.

My contributions in this research include collected, processed and analysed LiDAR data, design the overall study, interpreted results, prepared the manuscript, and coordinated revisions of the manuscript. Iain Woodhouse supervising the work, assisted with interpretation of the results and reviewed the manuscript. Hamdan Omar, assisted in designing the fieldwork experiment, assisted in data collection, and provide access to compilation of primary and secondary dataset and reviewed the manuscript and Carlos Silva reviewed the manuscript

5.1 Introduction

Individual trees information is required in a variety of forest-related activities, such as biodiversity assessment, selective cuts, silviculture treatment and improve global vegetation modelling (Lichstein et al. 2010). Advances in image sensing and Light Detection and Ranging (LiDAR) technologies make individual tree based analysis is feasible. Individual tree crown (ITC) serve as a basic unit for many useful activities such as volume or biomass estimation, species identification and gap analysis which has driven the development of various methods of ITC delineation (Hu et al. 2014). Airborne LiDAR has become the dominant technology in providing highly detailed spatial information of trees such as the number, location, spacing and size distribution of individual trees that may improve biomass estimation (Duncanson et al. 2014). LiDAR data also helps to improve forest parameter extraction at varying spatial resolutions and provide more ecologically meaningful structural description of a forest.

There are various algorithms or methods that have been proposed to detect or identify individual trees or tree crowns using airborne LiDAR data (Ferraz et al. 2016; Lee et al. 2017; Wan-Mohd-Jaafar et al. 2017). A good overview of the field is contained in Hyyppä et al. (2008). Most of the methods are based on raster-based approaches utilizing the canopy height model (CHM) (Chen et al. 2006; Silva et al. 2016). The canopy height model (CHM) derived from LiDAR data has been commonly used to extract segments of the individual tree crowns for forest inventory and sustainable management. However, tree crowns, tree clusters and branches have similar shapes and overlap between one and another, which causes current individual tree crown delineation methods to work less effectively on closed canopy or deciduous forest (Hu et al. 2014). In a typical CHM-based scenario, a local maximum filter is used to detect treetops and then the individual tree crowns are delineated with a marker-controlled watershed segmentation scheme or a pouring algorithm (Popescu et al 2003; Tiede et al. 2005; Zhen et al. 2014; Koch et al 2006 & Reitberger et al. 2009). Region growing is another method of image segmentation, the algorithm of region growing method

attempt to form segments of adjacent pixels that are similar in terms of gray value or texture. This algorithm tends to create segments with constant slope rather than with constant elevation when applied to the detection of planar regions in laser data (Gorte 2002). Region growing, as a category of simple region-based segmentation methods, can be adopted for segmentation of 3D point cloud generated by the LiDAR (Vo et al. 2015) and it is the most common approach to segmenting trees in original LiDAR data (Lee et al. 2010; Li et al. 2012; Alexander et al. 2009).

Original method of segmentation based on the use of watershed lines has been developed in the framework of mathematical morphology. This method, appear to be close to the growing-region methods, leads to a general methodology of segmentation and has been successfully applied in many different situations (Beucher & Meyer 1993). Gradient techniques in mathematical morphology watershed algorithm is one of the popular detection techniques. The difference between watershed-based delineation and local maxima is, watershed-based techniques lies on the improvement for crown geometries and function on the inverted CHMs by segmenting neighbouring crowns along lines of local minima.

For methods based on CHM-oriented tree crown delineations, in order to delineate varied-size tree crowns, CHM is typically smoothed to suppress spurious tree tops and local maximum filtering are used to detect the tree tops by using local maximum filtering with fixed or variable-size window (Persson et al. 2002). For a dense tropical forest, it is difficult to determine an optimal filter size for the local maximum filtering to retain multi-scale tree crowns because the allometric relationship between tree height and crown width is not always present and sometime so weak (Brandtberg 2002). Multi-scale analysis techniques have been used to overcome the issue caused by multi-scales of tree crowns in crown delineation (Liu et al. 2016). CHM-oriented tree crown delineations approaches have been centred at single scale and multi-scale analysis. However, multi-scale delineation methods are more able to account for trees of different sizes and have the similarity with the general practise of the human vision

system (Wang 2010). Plus, Jing et al. (2012) have shown that multi-scale methods performed better than those employing only one particular size or scale.

In the context of ITC delineation, different strategies or techniques have been used to achieve the multi-scale analysis. Falkowski et al. 2006 used a varying size of 2-D wavelets to filter a CHM image and the locations of multi-scale tree crowns were then detected from these filtered CHM images. Wolf and Heipke (2007), used a series of Laplacian of Gaussian filters to process a digital surface model (DSM); each of the filtered image was then segmented using a marker-controlled segmentation method and the crown segments were finally generated from the resulting multiple segmentation maps and refined. A series of 3D crown models with similar shapes and varied sizes was employed by Holmgren et al. (2010) to detect tree crowns based on the correlation between the CHM and the models. In all aforementioned methods, it was completely assumed that the Gaussian function, Mexican hat wavelets or 3D crown models resembled the 3D geometric shapes of multi-scale tree crowns and the correlation between the CHM and each model could drive the discrimination between tree crowns of different shapes. However, the scale level used in those methods were normally set manually or determined through trial and error and despite apparent success, these methods typically yielded noticeable errors of omission and commission especially in mixed wood, closed canopy or deciduous forest (Vauhkonen et al. 2010).

With the advancement of LiDAR systems to increase point density, several studies have attempted to segment individual trees directly using LiDAR 3D point clouds data. Extracting individual trees directly from the original LiDAR data can also resort to clustering-based methods. One of the most popular clustering methods is the k-means algorithm, which aims to partition n observations into k clusters where each observation belongs to the cluster with the nearest mean by trying to minimize the overall sum of Euclidean distances of the points in feature space to their cluster centroids (Lloyd, 1982; Morsdorf et al., 2003; Gupta et al., 2010). Lee et al. (2010) developed an adaptive clustering method to delineate individual trees in a managed

pine forest from 3-D LiDAR data, which required a large number of training samples for a supervised learning. Li et al. (2012) adopted a top-to-bottom region growing approach that segmented individual trees sequentially from the tallest to the shortest based on 3-D structures of tree crowns captured by LiDAR data. In recent years, mean shift-based clustering techniques have been applied in individual tree delineation using LiDAR data (Ferraz et al. 2010; Ferraz et al. 2012; Ferraz et al. 2016; Hu & Xie 2016; Amiri et al. 2016). Mean shift clustering is considered the most recent and sophisticated approach for individual tree crown delineation and has found much interest in the image processing and computer vision community (Comaniciu & Meer 2002). The main drawbacks of these methods working with direct 3D LiDAR point clouds is that it requires more computation power as it is working with a huge number of data points rather than with raster images. Plus, working with massive 3D point clouds is challenging to extract useful crown features from the LiDAR returns generated by various objects in a forest scene especially when the methods do not depend on any geometric model assumptions as algorithm like mean shift is built on probabilistic intuitions.

In this chapter, to take advantage of both the simplicity of the CHM-oriented methods and detailed 3D structures of tree crowns, an innovative framework to improve the accuracy and efficiency of ITC delineation from LiDAR data was developed and act as the main contribution of this study. In addition, the methods proposed will automatically determine the dominant tree crown sizes in each of forest site and to detect the pattern of clustering patter of species, which will be studied in the next Chapter 6.

5.2 Study areas

Details of the study areas were explained in detail in section 3.1. The CHM generated explained in detail in section 4.3.1 was smoothed with a 3 x 3 Gaussian low pass filter to effectively eliminate noise as done in Morsdorf et al. (2004). The generation of CHM as a derivation of the difference between DSM and DTM for PFR and FRIM site are shown in Figure 5.1 and 5.2.

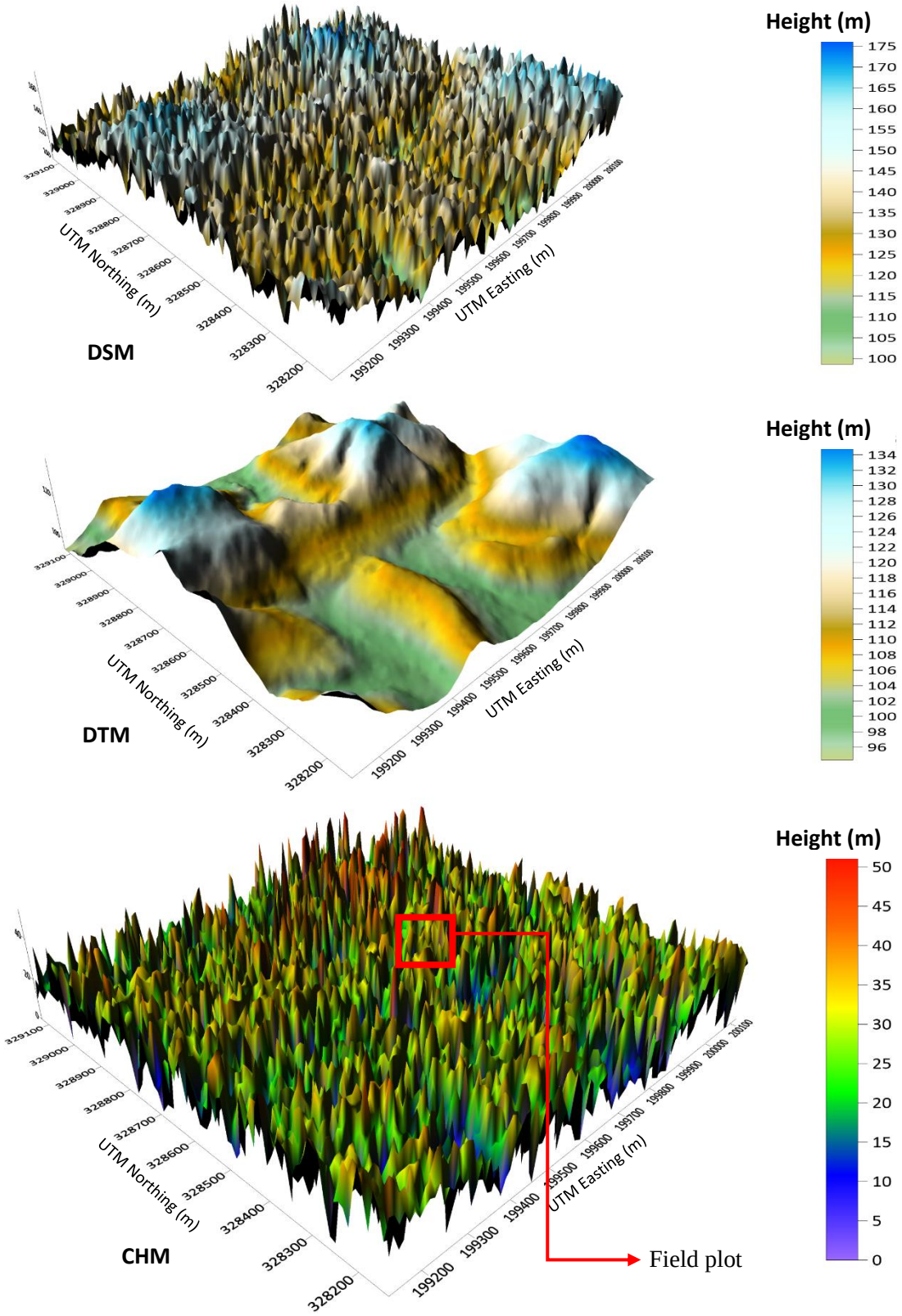


Figure 5.1. Derivation of CHM as the difference between DSM and DTM for PFR site.

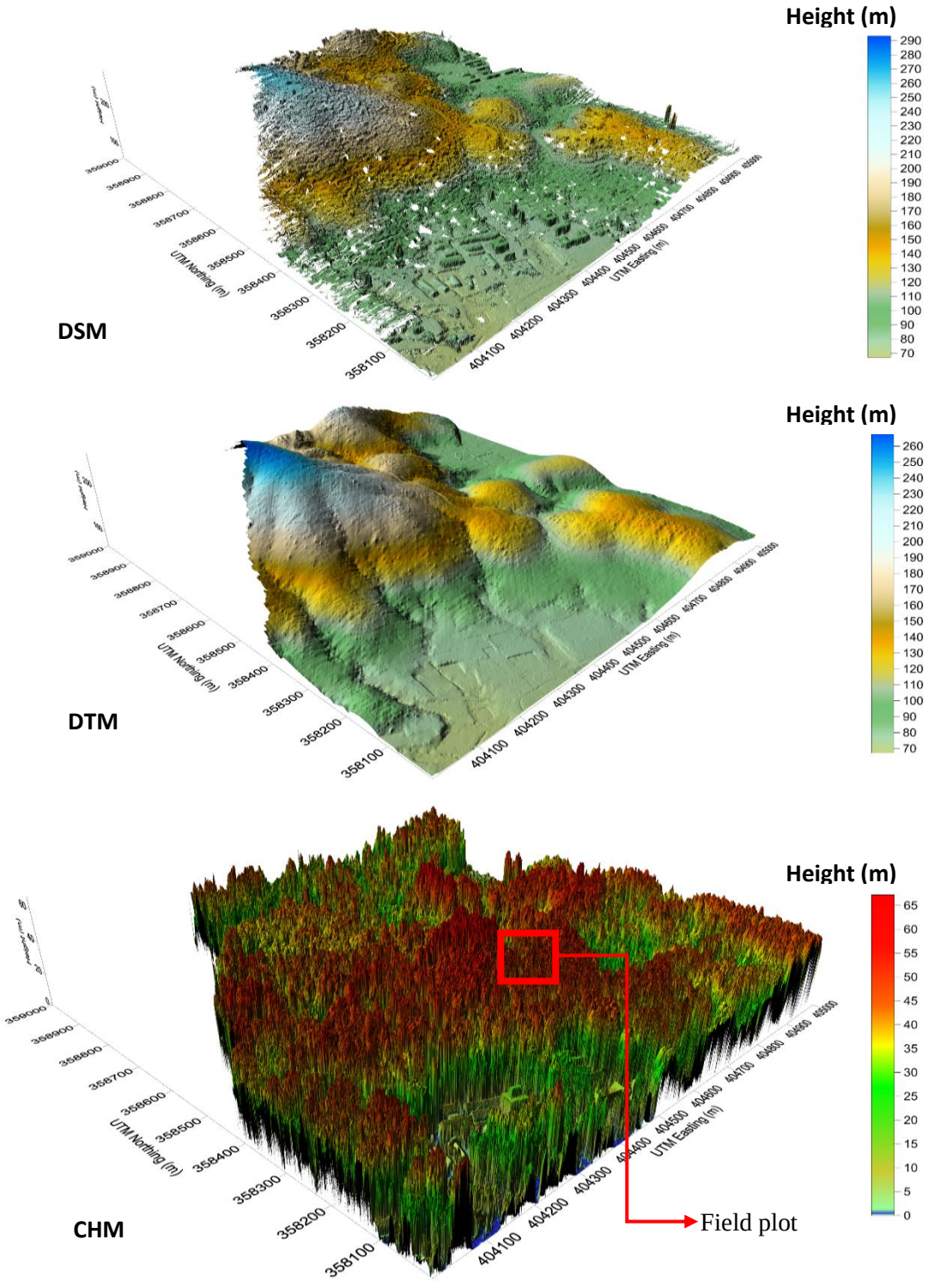


Figure 5.2. Derivation of CHM as the difference between DSM and DTM for FRIM site

5.3 Methodology

In this research chapter, a framework to improve accuracies of ITC by adapting raster based segmentation and 3D point cloud data was developed. The framework consisted of five steps (Figure 5.3) : (1) determine the crown size from the CHM image through automatic procedure using semi-variogram statistics and morphological analysis; (2) generate the initial tree segments from the CHM image based on the determined crown size (scale levels) using watershed segmentation based on mathematical morphology; (3) evaluate the initial tree segments and identify the problematic segments for refinement by a set of rules determined prior to the knowledge of study sites; (4) determine the number of trees detected based on the 3D LiDAR points in each of the identified segments; and (5) refine the problematic segments by using modified distance based algorithm developed by Li et al. (2012).

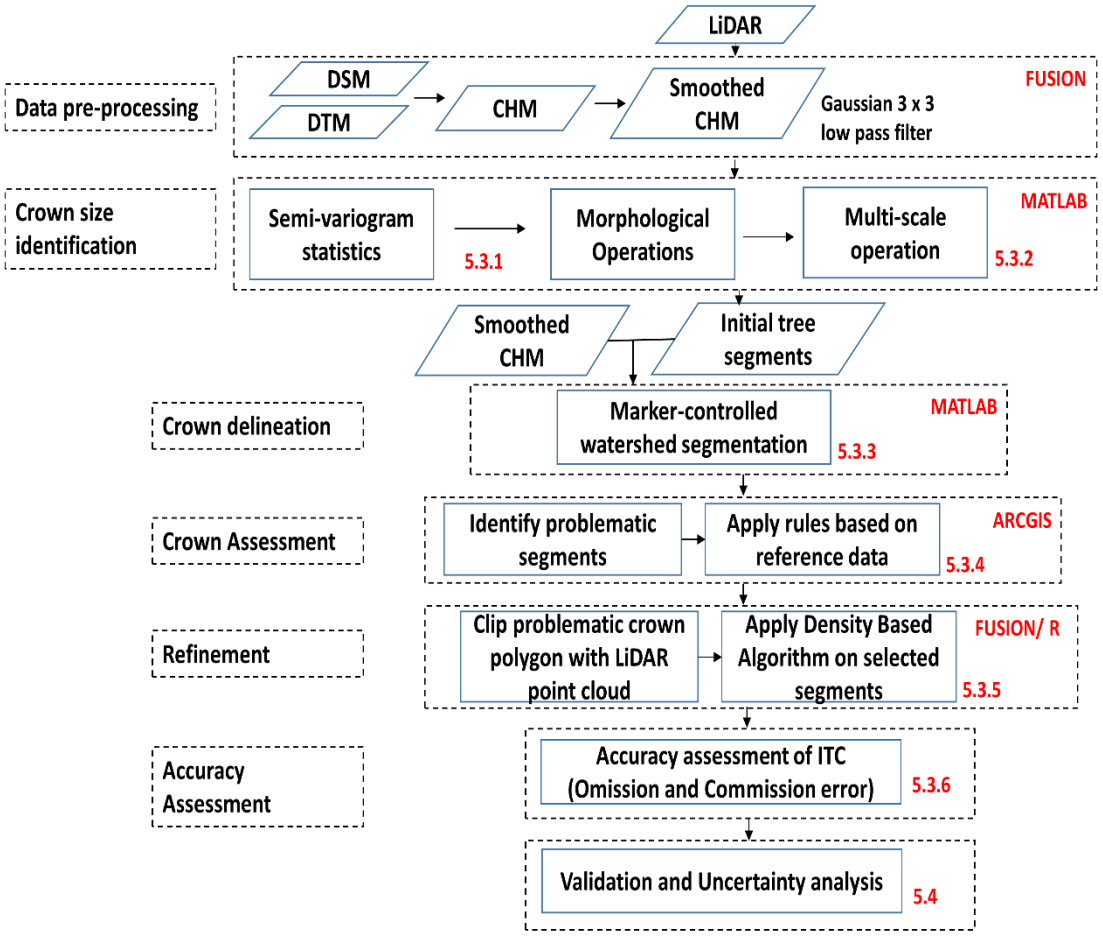


Figure 5.3. Individual tree crown delineation flow diagram

5.3.1 Spatial Structures Characterization

Morphological opening operations with appropriate structuring elements (SE) can separate different sized objects within a grayscale image based on sizes and shapes (Soille 1999). SE is a matrix consisting of only 0's (erosion) and 1's (dilation) that can have any shape and size. From the opened image result, objects that completely cover the SE are truncated and retained, while others are sifted out. If opening operations with disk SEs of a series of sizes are applied to the CHM image of a forest scene, different-sized tree crown can be potentially being separated. By looking into the core of the basic principle of morphology technique, the 3D radiometric shape of a tree crown can be considered as a half ellipsoid (Figure 5.4). The half ellipsoid represents the coarse and fine structures of this crown, respectively. If this tree crown is sliced into many layers from top to bottom (Figure 5.4 a1-c1), a disk can be fitted into each resulting slice. The height values decrease continuously from the treetop to crown the boundary. From a 3-D perspective, the algorithm vertically scanned the CHM from top to bottom by a horizontal plane. The diameter of the largest disk lying at the bottom slice represents the size of the half ellipsoid and size of the tree crown. As a result, the size of the tree crown can be measured using a morphological opening operation with a disk structuring element (SE) (Soille 1999).

This process can better be explained through the graphical illustration as shown in Figure 5.4. In the process, the higher individual tree (tree B) produced a cross-section earlier (Figure 5.4-a2) than the shorter tree (tree A). Figure 5.4-b2,c2 indicated that the two cross-sections dilated with the decrease of scanning height. The cross sections of tree A and B appeared at previous step must be contained in the subsequent cross-section. Cross-sections of individual tree crown regions that do not contact with others typically appear circular, as shown in Figure 5.4-a2,b2 and the cross-sections produced by overlapped trees often appear irregular shape, like Figure 5.4-c2

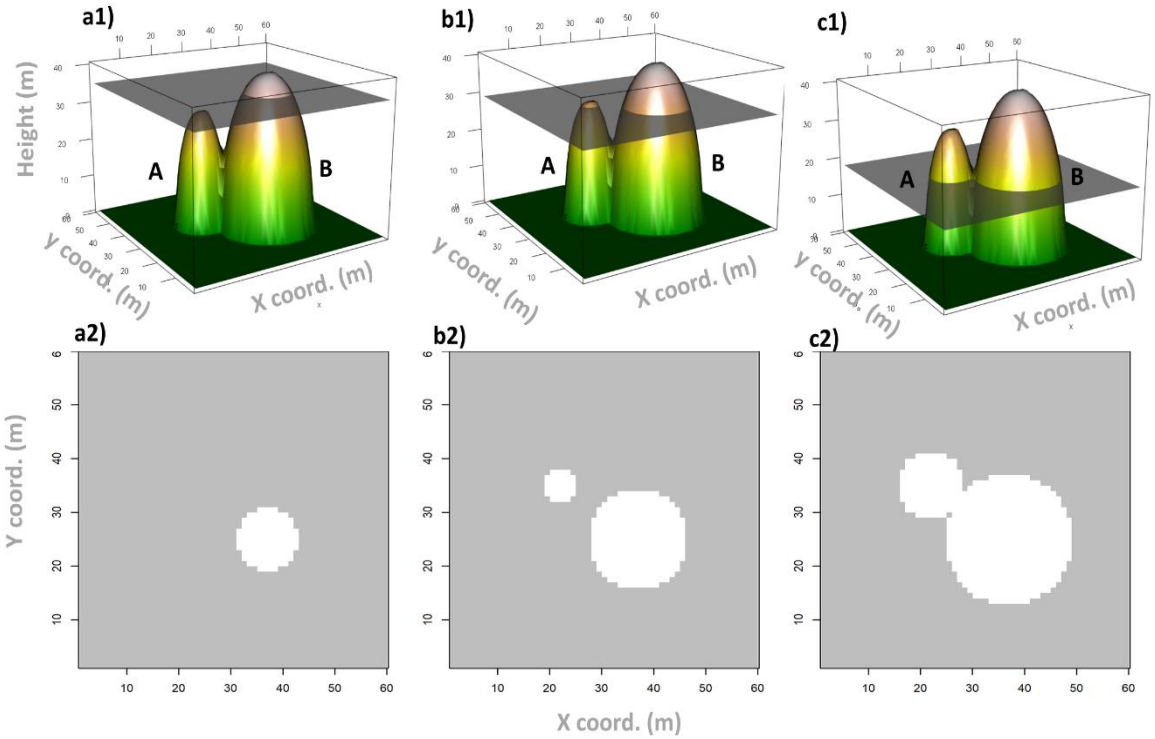


Figure 5.4. The illustration of separating different-size crowns using the disk SE based on a CHM containing tree A and tree B. (a1-c1) represent the scanning moving window process on CHM; (a2-c2) are the corresponding cross-sections produced from (a1) to (c1), respectively.

The CHM generated in this study is 1 m resolution. To explore the potential sizes of tree crowns, a series of disk SEs with diameters 1 to 19 pixels was used in the morphological opening operation on a CHM image, and a series of opened CHM images was generated. These opened CHM images are marked as OC_i, i= 1, 3,…, 25 where *i* is the diameter of the SE used to generate OC_i. For any two consecutive opened images, OC_i and $OC_{i+2,}$ where subscripts indicate the corresponding SE diameters, their difference image (OC_{i+2} - OC_i) was computed and its mean value was calculated. All the mean values of the difference images of the series of opened images for PFR and FRIM are shown in Figure 5.4. As seen in Figure 5.4, there are several local minima. A local minimum occurs when there are significant differences in object sizes between two adjacent opened images. Therefore, these mean values reveal the

dominant sizes of the objects and their size range captured in the image. The size range of the objects in site PFR and FRIM is wide and multiple dominant size groups exist. For 1 meter = 1 pixel size, the size ranges: 1-2 pixels (1.0 – 2.0 m), 3-5 pixels (3.0 – 5.0 m), 7-11 pixels (7.0 – 11.0 m), 13 – 15 pixels (13.0 – 15.0 m) and 17- 19 pixels (17.0 – 19.0 m). The first group was considered as the sizes of branches and the four other groups were taken as the sizes of small, medium and large tree crowns and tree clusters respectively. By taking the minimum value within each tree crown size group as group representative, three dominant tree crown sizes of 1,3 and 7 pixels for PFR site and 3, 7 and 13 pixels for FRIM site were obtained, and hereafter, they are referred as the small, medium and large tree crown level respectively (Figure 5.5 and 5.6). Further analysis is needed to isolate the tree crowns.

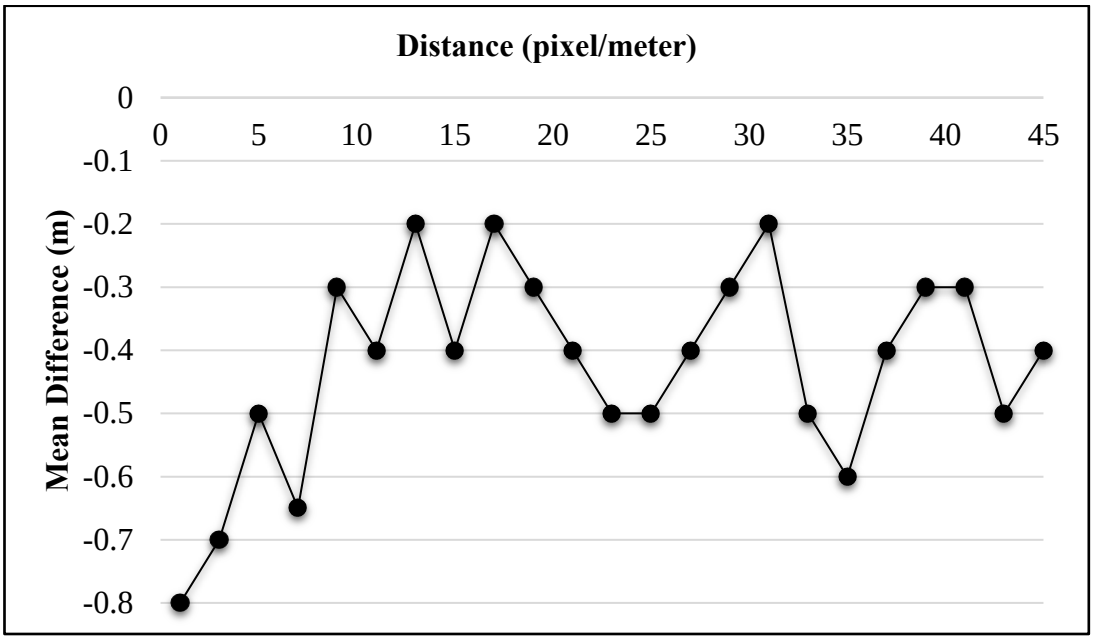


Figure 5.5. Difference of the mean values of the series of opened images for PFR site obtained using a disk SE of a diameter *i* pixels.

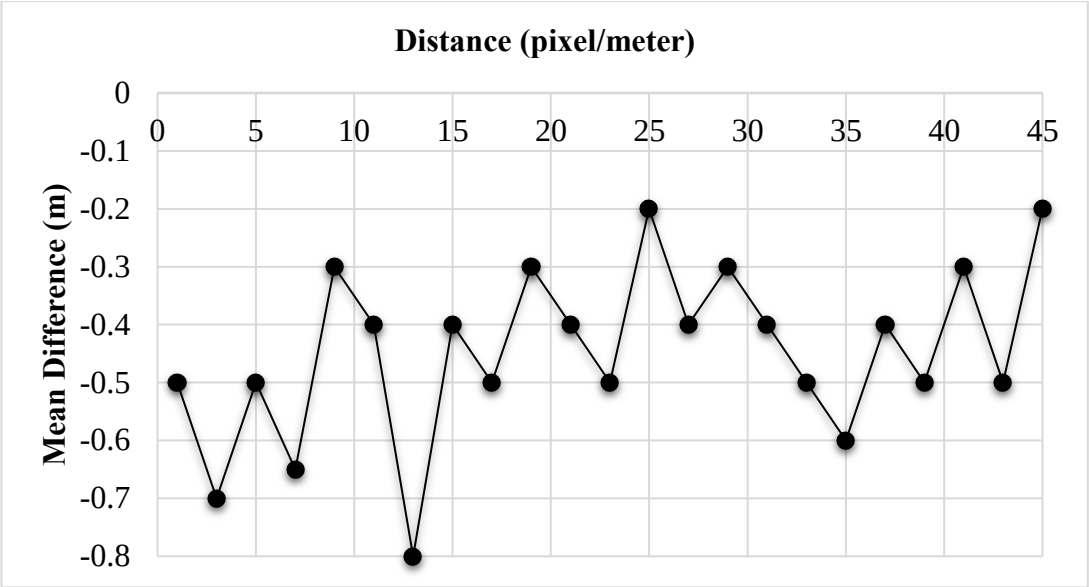


Figure 5.6. Difference of the mean values of the series of opened images for FRIM site obtained using a disk SE of a diameter *i* pixels.

Semi-variogram statistics was used to find the range of crown sizes in the CHM image (Clark, 1979). A number of studies have successfully used semi-variogram to characterize the spatial structures of observed surface properties (St-Onge & Cavayas, 1995; Garrigues et al. 2006). The calculated semi-variogram for the scene in CHM PFR and CHM FRIM site are shown in (Figure 5.10). This semi-variogram measured the squared difference in canopy height between pairs of pixels with a given distance apart and thus indicated the spatial variability of the observed canopy height. Starting from 0, the semi-variogram value increased with the lag distance between the paired pixels with a varied rate as manifested by a "clear break" in the first-derivative of the semi-variogram occurred and this was interpreted as the minimum size of the dominant crowns in the scene. The distance where the semi-variogram reached a plateau was considered as the maximum crown size. The rationale from the author interpretation detailed as follows:

1. When a pair of pixel were separated by a very short lag distance, it was likely that they were either from the same crown or the same between-crown gap, and thus the difference in canopy height between them was small. The height difference gradually increased when the pixels moved away from each other but still belonged to the same crown or gap.

2. When the lag distance increased to a certain point where most paired pixels were either, one from a crown and the other from the immediate surrounding background or from two close proximity different crowns, the difference in height between them would be large and as a result a sharp rise in the semi-variogram and a clear break in its first-derivative were expected.
3. When the distance between a pair of pixels were beyond the maximum crown size, it was likely that they belonged to different canopy elements and their heights were independent of one another. Therefore, the variogram values were roughly constant beyond this distance.

Combining the range of crown sizes concluded from the semi-variogram and size groups determined by the morphological analysis, it can be concluded that tree crowns in PFR site were dominated by three size groups referred as small, medium and large tree crown: 1-2 pixels (1.0 – 2.0 m), 3-5 pixels (3.0 – 5.0 m) and 7-11 pixels (7.0 – 11.0 m) while FRIM site were dominated by 3-5 pixels (3.0 – 5.0 m), 7-11 pixels (7.0 – 11.0 m) and 13 – 15 pixels (13.0 – 15.0 m). Since the selected tree crown level was the minimum value within a group of crown sizes, objects with larger sizes within the group would be retained via filtering or morphological opening operation based on the minimum value.

5.3.2 Multi-scale Operations

The principal of this method is the identification of tree tops for marker-controlled watershed segmentation. The algorithm was designed based on the following assumptions and observations; (i) the upper part of a crown or an entire tree crown in a CHM image can be generalized as a half-ellipsoid with branches around the boundary, which represent coarse and fine structures of the crown and, (ii) the horizontal cross-section of a crown at a specific height contains its tree top and also indicates its horizontal extent at that height.

By applying the morphological opening operations with a disk SE of a diameter *d* to a CHM image, tree crowns smaller than the disk SE are removed, while others are truncated horizontally. The tree tops of the truncated crowns are equivalent to their horizontal cross-sections at certain heights. The cross-sections, which are the local maxima in the opened CHM image, are good indicators of the position and horizontal extent of tree crowns at the scale level defined by the SE. In the forest scene, trees vary in size and some branches resemble individual trees. Multiple opening operations with different-sized disk SEs are needed to sort all of the tree crowns, leading to multiple layers of cross-sections of tree crowns. Since in a forest some branches and tree clumps have similar sizes to individual tree crowns, and effective method is needed to merge different layers of cross-sections together to generate a layer of markers of tree crowns. A tree crown in a CHM usually appears as circular, whereas a tree cluster, as a combination of several crowns, is less circular (Jing et al. 2012). Circularity (c) threshold of a segment (Wolf & Heipke, 2007) can be calculated by:

$$c = A/\pi r^2 \ . \tag{5.3}$$

Where A is the area of the segment and r is the largest distance between the centroid and border of the segment. In this study, the author has computed height (percentile height), crown length (CL), crown ratio (CRatio), crown based height (CBH), crown radius (CRad), crown volume (CV), crown projection area (CPA) and crown surface area (CSA) from the LiDAR point clouds of individual tree by using the CrownMetrics function from the rLiDAR package (Silva et al. 2015). This value was adopted to calculate the circularity threshold of a segment. The circularity threshold was set to 0.85 in this study because the typical circularity values of tree crowns lie above 0.85 (Wolf & Heipke, 2007). The merging cross-sections of tree crowns method from Hu et al. 2013 was adopted in this study and it can be summarized as follows:

1. Remove cross-sections with circularity less than a threshold to eliminate tree clusters. As the threshold was set at 0.85 in this study, based on the observation made in Wolf and Heipke (2007) that the circularity of tree crowns was typically above 0.85

2. Logic 'OR' operation used to combine the cross-sections on both layers
3. Refine the merged cross-sections. This is done by removing those with circularity less than the threshold.

The resulting cross-sections indicated the positions of tree crowns in the scene. Identified positions were then used as a marker for watershed segmentation.

5.3.3 Marker-controlled Watershed Segmentation

In this study, the commonly used marker controlled watershed segmentation was used to segment the CHMs at different scales separately (Meyer & beucher 1990; Vincent & Soille 1991). CHM was generated by using *canopymodel* function using FUSION software (McGaughey 2014) as explained in section 4.3.1. This raw CHM is then smoothed using Gaussian filter. The filtering kernel of Gaussian filter is detail as follow:

$$g(x, y, \sigma) = \frac{1}{2\pi\sigma} e^{-(\frac{x^2+y^2}{2\sigma^2})} . \quad (5.1)$$

Where x and y are the distant to the centre of smoothing nuclear. σ is used to adjust the parameters of the Gaussian function and a 3 x 3 filtering kernel of Gaussian smoothing algorithm was used (Figure 5.7).

g (-1,1, σ)	g(0,1,σ)	g(1,1,σ)
g(-1,0,σ)	g(0,0,σ)	g(1,0,σ)
g(-1,-1,σ)	g(0,-1,σ)	g(1,-1,σ)

Figure 5.7. 3 x 3 filtering kernel of Gaussian smoothing algorithm

The CHM was smoothed by using this moving filtering kernel. For pixel; p(x, y) located in the centre of filtering kernel, its DN value is computed as:

$$p(X,Y) = \sum_{i=-\frac{d}{2}+1.0}^{\frac{d}{2}-1.0} \sum_{j=-\frac{d}{2}+1.0}^{\frac{d}{2}-1.0} p(X+i, Y+j) \text{ x } g(x+i, y+j, \sigma)\,. \quad (5.2)$$

X and Y are coordinates of the pixel in the image, *d* (tree crown size of *d* pixels) is the size of the filtering template which is an odd number such as 3, 5, 7, *i* and *j* are the distant to the studied pixel. This was design to take account of various 3D radiometric shapes of tree crowns, similar to the ITC delineation methods introduced in Persson et al. 2002; Brandtberg et al. 2003 and Falkoswski et al. 2006. Tree crowns with a similar shape and size could be enhanced and smaller objects could be effectively suppressed when a Gaussian filter was applied to a grayscale image. When the filter was applied to the CHM, tree crowns with shapes and sizes similar to the filter were enhanced, while smaller object were suppressed. Since the tree crown level was the minimum value within a tree crown size group, objects with larger sizes within the group would be retained. The three crown levels previously determined led to three Gaussian filters and thus three filtered CHMs with each containing object with similar sizes and shapes.

The watershed simulates the immersion from markers to determine the flooded basins. The key step in the marker-controlled watershed segmentation is the determination of markers. The local maxima in the CHM at each given scale were employed as markers. With the detected local maxima in the CHM at each given scale serving as markers, the marker-controlled watershed segmentation was applied to the CHM image by using the functions implemented in the Matlab (version R2016a). After removal of non-crown areas such as buildings and bushes or grass (marked as anything below than 2 metres), which were previously delineated by a threshold method on the CHM, the segmentation maps at the multiple scale level were generated for both forest site (Figure 5.8 and Figure 5.9).

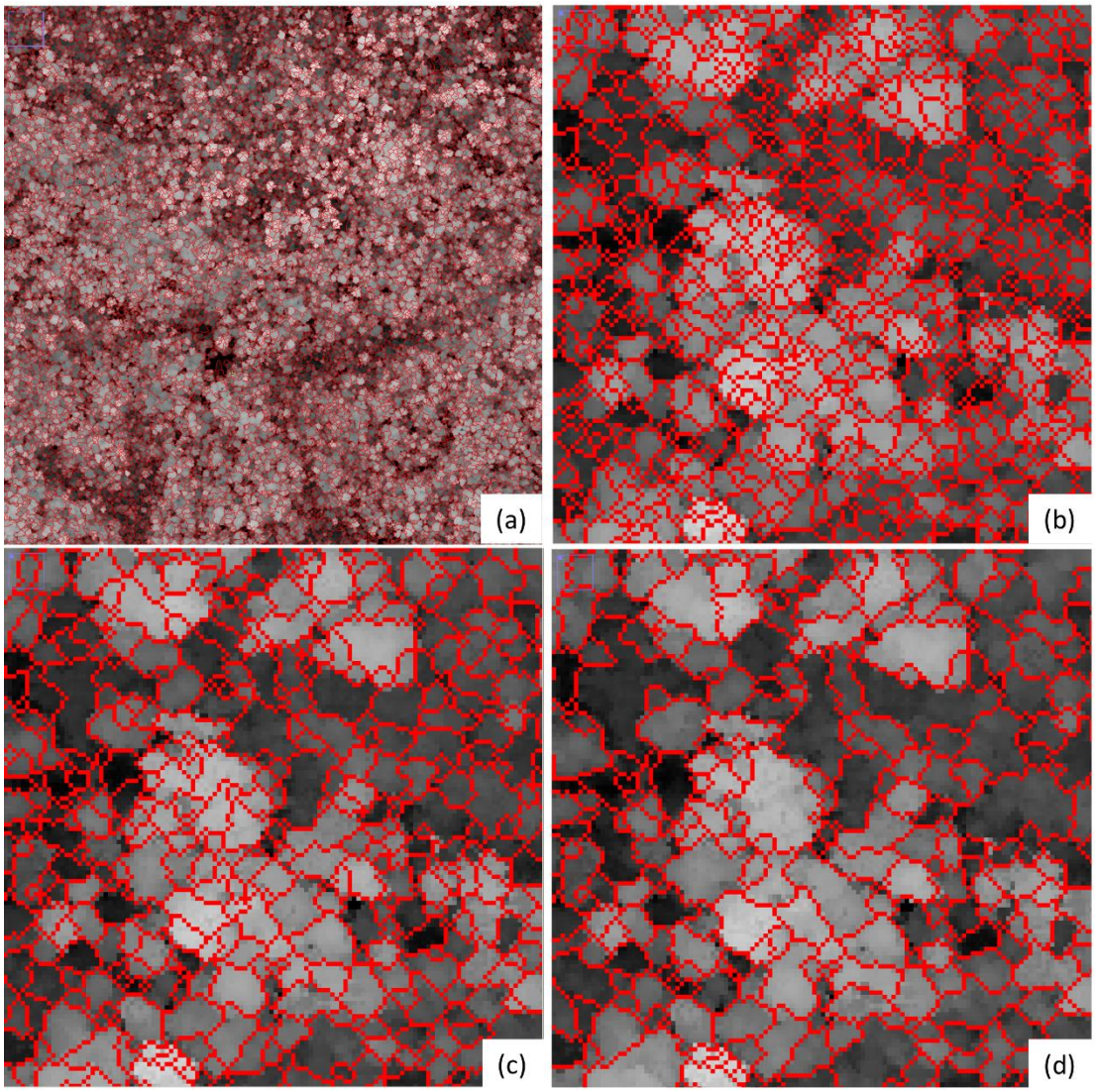


Figure 5.8. The crown segmentation map of PFR site (a) crown segmentation map over 100 ha and the segmentation at (b) small, (c) medium and (d) large tree crown levels with the corresponding filtered images as background.

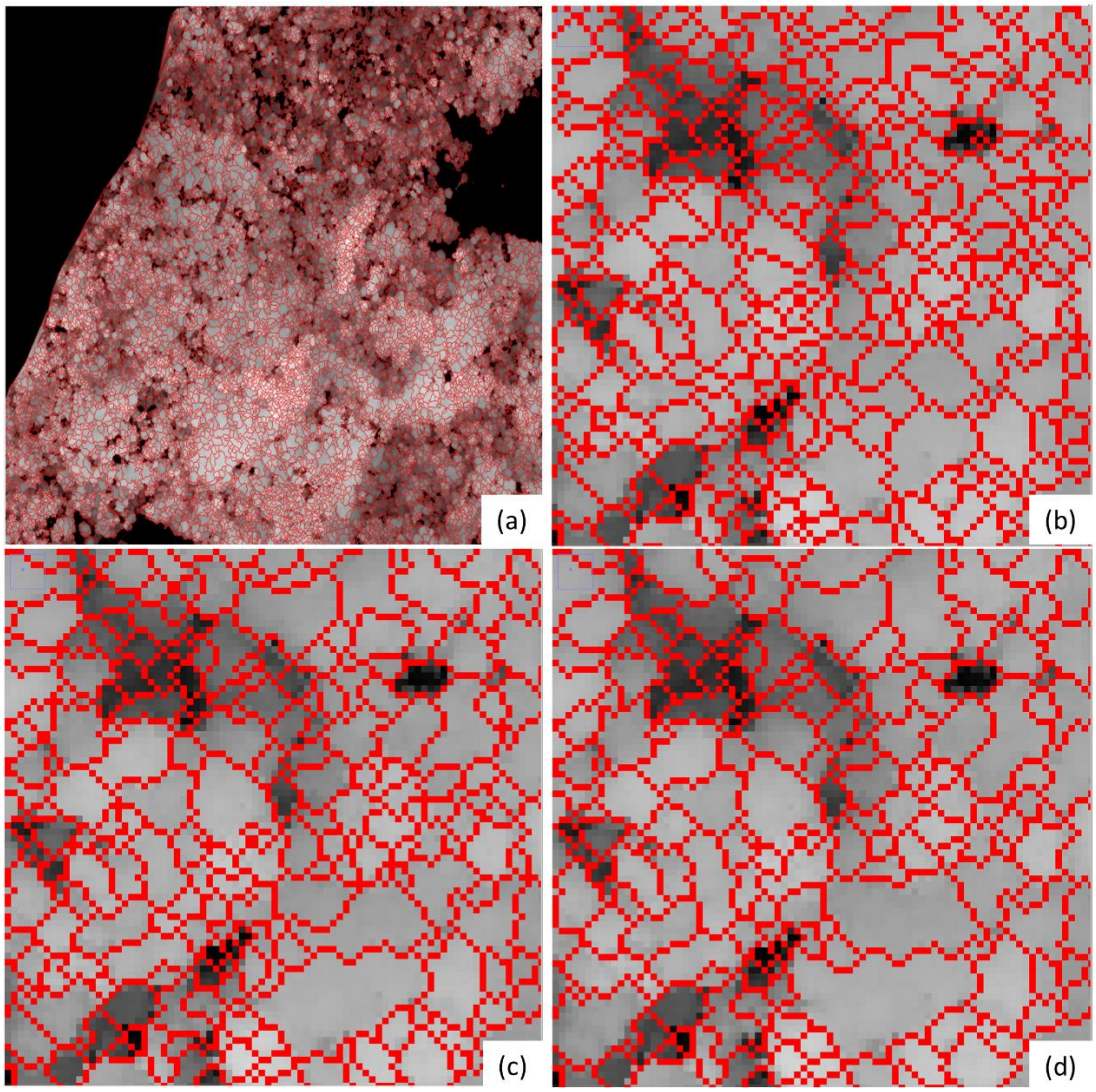


Figure 5.9. The crown segmentation map of FRIM site (a) crown segmentation map over 100 ha and the segmentation at (b) small, (c) medium and (d) large tree crown levels with the corresponding filtered images as background.

5.3.4 Identification of Problematic Tree Segments

The preliminary watershed may still represent a single tree or a cluster of trees. To separate neighbouring trees forming a tree group and to improve the accuracy of the individual tree positions, a few rules was set based on the crown size detected and comparison with the reference trees. This second evaluation will be conducted on the selected trees that is motivated by the observation from the centre segment of Figure

5.10 contain two trees, which cannot be separated by the watershed segmentation, whereas the stems of at least two of the trees is known exist in the selected polygon based on the tree reference (as discussed in Chapter 4) and based on the rules criteria of the crown size in the study area. A second segmentation was run on trees that has been identified the need to be segmented again.

A set of rules was used to identify segments for further refinement. To prove the concept, only general knowledge of tree crowns was applied. As for the first general information of the crown segments that was collected during the fieldwork:

- Trees with crown size greater than 11 m (>11m) for PFR site and trees with crown size greater than 17 m (>17m) for FRIM site identified as not a single tree but a group of trees that need to be segmented again because the largest crown size detected at PFR site was 10.3 m and 16 m at FRIM site (reason for the selection crown size criteria).

The following conditions were flagged for further examination if the following rules were met:

- The circularity of a small segment was large or the circularity of a large segment was small. The criterions assume that if a segment is large with a low circularity index, it is likely a tree cluster. The circularity threshold was set to 0.85 to detect the large segments

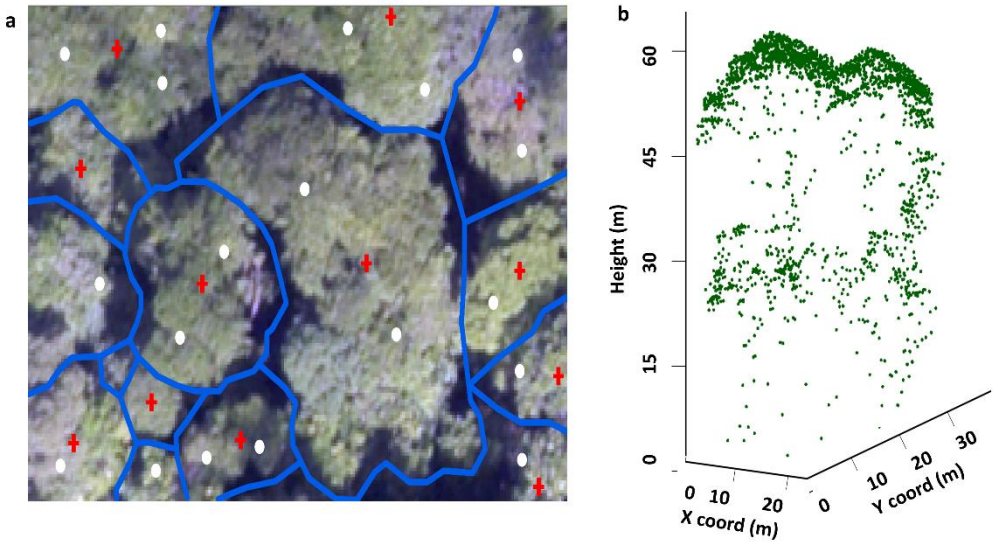


Figure 5.10. Digital orthophoto used to visualize the results (a) Orthophoto with watershed segments (blue), field reference tree (white) and local maxima (red cross) and, (b) LiDAR point clouds for the segment

5.3.5 Refinement of the final tree crown delineation using Distance-based Algorithm

Usually, in order to segment 3D scenes, approaches inspired from 2D are used. These methods often involve a pre-processing step in order to find key points in the scene before applying to a more robust 3D approach to the key points to find the final segments (Hadji & Nabelek n.d.). Distance-based tree algorithm (Li et al. 2012) were applied to any segment identified in section 5.2.3 to section 5.2.4 with more than one tree or with no-tree respectively using *lidR* R package (R Development Core Team, 2016). The proposed segmentation algorithm isolates trees individually and sequentially from the point cloud, from the tallest to the shortest by adopting top-to-bottom approach to classify the points. If more than one was identified in a given segment (e.g., *n* trees), distance-based algorithm was used to separate the *n* tree points in this segment. The algorithm exploits the spacing between the tops of trees to identify and group points into a single individual tree. The main uncertainty for tree segmentation in this algorithm was mainly derived from the spacing threshold. A higher threshold can result in under-segmentation whereas a smaller threshold can result in over-segmentations. For this step, the author used an adaptive threshold (*dt*) to improve the segmentation accuracy: *dt* = 1.5 and 1.0 meter when tree heights were $\geq$ 15 and $<$15 meter, respectively. This spacing threshold was chosen based on the author knowledge of the forest study areas, plus incorporated with the crown rules as discussed in section 5.3.3. It is assumed that the tree spacing at the upper level is > 1 meter and the taller trees have larger crown sizes results to larger spacing. A minimum spacing rule was not used as it did not improve the accuracy of the classification.

5.3.6 Determination of numbers of trees detected in a segment based on 3D LIDAR points

To quantitatively assess the accuracy of the segmentation results, we compare the identified trees with the reference trees in 16 testing subplots (8 subplots per forest site). As mentioned in Li et al. (2012), the measures of perfect segmentation, under segmentation and over-segmentation can be indicated by true positive (TP), false negative (FN) and false positive (FP), respectively. We can evaluate the detection accuracy in terms of "recall", which indicates the tree detection rate and "precision" that indicates the correctness of detected trees. The "recall" (r) and "precision" (p) are defines as follows (Goutte & Gaussier, 2005; Sokolova et al. 2006):

$$r = \frac{TP}{TP+FN'} \quad . \tag{5.4}$$

$$p = \frac{TP}{TP+FP'} \quad . \tag{5.5}$$

$$F = 2 * \frac{r * p}{r + p} \quad . \tag{5.6}$$

Recall indicates the tree detection rate and inversely related to omission error. Precision indicates the correctness of the detected trees and inversely related to commission error and F-score is the overall accuracy, taking both the commissions and omission errors into consideration and used to represent the related mean of recall and precision. A higher F-score indicates a higher *r* and *p* hence, lower commission and omission errors (Li et al. 2012). Recall, precision and F-score ranges from 0 to 1. If all of the trees are correctly segmented, the r and p values are one which resulting in F being equal to one.

5.4 Results and Analysis

5.4.1 Evaluation of morphological watershed segmentation

As stated in section 5.3.1, the dominant crown sizes in PFR size were 1, 3 and 7 pixels, while for FRIM site the dominant crown sizes were 3, 7 and 13 pixels as determined from the semi-variogram and its first derivative shown in Figure 5.11 and 5.12. As mentioned earlier, the range position calculated by fitting a spherical model to the semi-variogram as can be shown from the figure. The position at the first "clear break" of the semi-variogram's first derivative indicated the minimum crown size. As a result, there were four groups of crown sizes. The smallest sizes of each group denoted 1, 3 and 7 pixels for PFR site and 3, 7 and 13 pixels for FRIM site were selected and used in watershed segmentation process. The cross sections mark as the opening of the size of disk SE will be using as the markers in the marker-controlled watershed segmentation to generate the tree segments. By observing the segment boundaries overlaid over the original CHM images, most of the crowns shows a decent result but some of the tree segments were detected as a group of trees or clusters based on the field reference and through raw vision. Based on the criteria as discussed in section 5.3.3, Figure 5.13 shows the sample of segments in point clouds that needed further refinement and the results after refinement by using the distance-based algorithm.

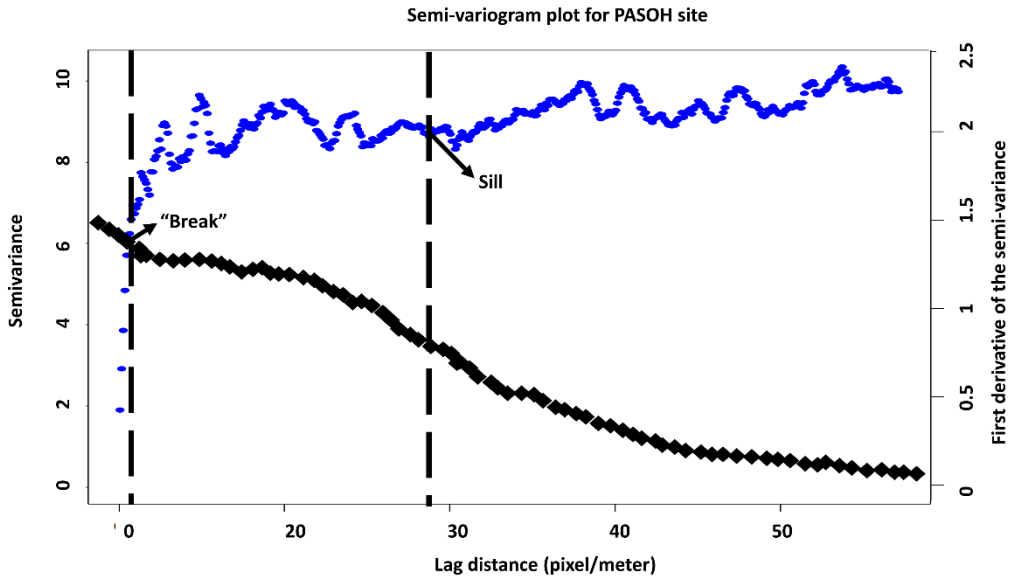


Figure 5.11. The semi-variogram (dotted line with blue markers) and its first derivative (diamond shape with black markers) of the CHM PFR site. The left vertical dashed line indicates where the break in the slopes of the semi-variogram occurs, and the right vertical dashed line indicates the range where the semi-variogram reach its sill.

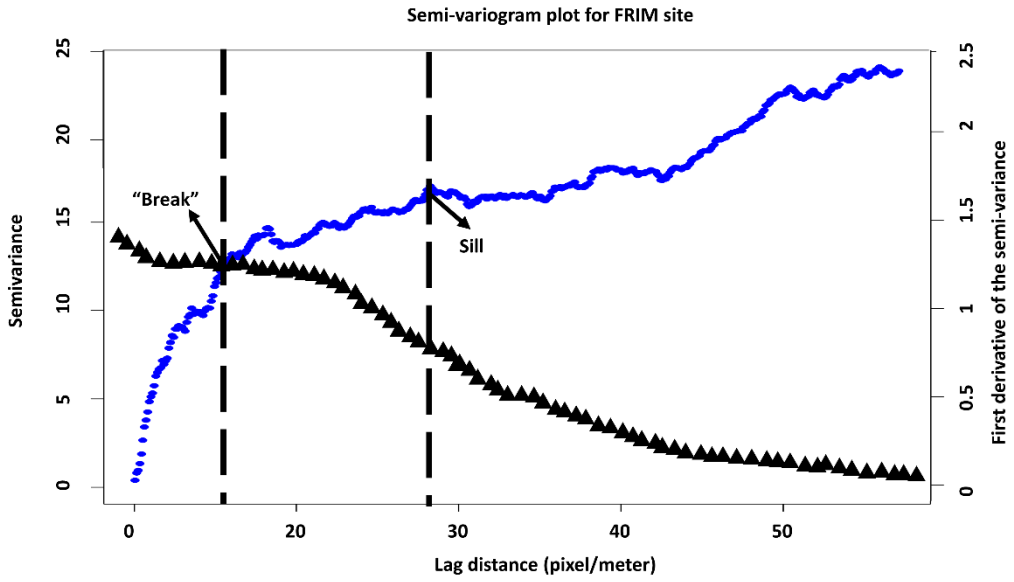


Figure 5.12. The semi-variogram (dotted line with blue markers) and its first derivative (triangle shape with black markers) of the CHM PFR site. The left vertical dashed line indicates where the break in the slopes of the semi-variogram occurs, and the right vertical dashed line indicates the range where the semi-variogram reach its sill.

Figure 5.13. Comparison of results between watershed delineation and result after refinement using DBA algorithm. The two trees on the right were segmented using watershed segmentation; same two trees on the left were refined using DBA algorithm. The refinement produced point cloud that better represent the bole and lower portion of the tree, while watershed segmentation delineated every point below the surface, including extraneous features that does not belong to the tree.

Accuracy statistics of the ITC delineation generated using the watershed segmentation with mathematical morphology are listed in Table 5.1 and 5.2 while table 5.3 and 5.4 shows the accuracy statistics of ITC delineation after refinement using the distance based algorithm.

5.4.2 Accuracy assessment of Individual tree detected

The accuracy assessment results for individual tree detection from the watershed segmentation results in the two forest site is shown in Table 1 and Table 2. For PFR site, the recall varies from 0.75 to 0.86, with the overall value of 0.81; the value of p varies from 0.36 to 0.91, with the overall value of 0.79; and the F-score, which considers both of these last 2 factors, varies from 0.53 to 0.87, with the overall value from all the plot of 0.80. While FRIM site, the recall varies from 0.78 to 1.0, with

the overall value of 0.85; the value of p varies from 0.38 to 0.75, with the overall value of 0.69; and the F-score, which considers both of these last 2 factors, varies from 0.52 to 0.86, with the overall value from all the plot of 0.76.

There are 105 reference trees from PFR site and 109 (80.95%) trees were detected and 104 reference trees from FRIM site and 127 (84.6%) trees were detected. In summary, the algorithm missed 20 (19 %) trees, and falsely detected 22(21 %) trees at PFR site, while at FRIM site, the segmentation algorithm missed 16 (15.4%) trees, and falsely detected 39(37.5%) trees, both with the same condition, overdetection outweighing underdetection (Table 5.1 and 5.2).

Distance-based algorithm approach was applied to the identified problematic segment to refine the crown results and the results were improved with 106(83.8%) trees were detected from the 105 reference trees at PFR site and 118 (88.46%) trees were detected out of the 104 numbers of reference trees at FRIM site. Distance based algorithm results improve the results where the algorithm only missed 17 trees (16.2%) and falsely detected 18 trees (17.14%) at PFR site, and it does improve the results to the FRIM site too with only 12 (11.53%) trees missed and 26(25%) falsely detected trees. Even though the refinement results still show overdetection condition, but it is slightly improving the results and it is considered good enough for one working in a dense forest.

Table 5.1: Accuracy assessment for tree segmentation in PFR site over 8 testing plots according to statistics parameters

	Number of Trees Detected							
Subplots	**LiDAR**	**Reference**	**FP**	**FN**	**TP**	**r**	**p**	**F**
A1	22	24	2	4	20	0.83	0.91	0.87
A2	20	23	2	5	18	0.78	0.90	0.84
A3	26	22	5	3	19	0.86	0.79	0.83
A4	7	8	1	2	6	0.75	0.86	0.80
B1	11	4	7	0	4	1.00	0.36	0.53
B2	7	8	1	2	6	0.75	0.86	0.80
B3	10	8	4	2	6	0.75	0.60	0.67
B4	6	8	0	2	6	0.75	1.00	0.86
Overall	**109**	**105**	**22**	**20**	**85**	**0.81**	**0.79**	**0.80**

FP: False positive; FN: False negative; TP: True positive; r: recall; p: precision and F: F-score

Table 5.2: Accuracy assessment for tree segmentation in FRIM site over 8 testing plots according to statistics parameters

	Number of Trees Detected							
Subplots	**LiDAR**	**Reference**	**FP**	**FN**	**TP**	**r**	**p**	**F**
A1	18	13	6	1	12	0.92	0.67	0.77
A2	21	17	7	3	14	0.82	0.67	0.74
A3	20	15	5	0	15	1.00	0.75	0.86
A4	23	18	7	2	16	0.89	0.70	0.78
B1	13	11	5	3	8	1.00	0.50	0.67
B2	14	12	4	2	10	0.83	0.71	0.77
B3	9	9	3	3	6	0.89	0.38	0.53
B4	9	9	2	2	7	0.78	0.39	0.52
Overall	**127**	**104**	**39**	**16**	**88**	**0.85**	**0.69**	**0.76**

FP: False positive; FN: False negative; TP: True positive; r: recall; p: precision and F: F-score

Table 5.3: Accuracy assessment for ITC after refinement using distance-based algorithm in PFR site according to statistics parameters

	Number of Trees Detected							
Subplots	**LiDAR**	**Reference**	**FP**	**FN**	**TP**	**r**	**p**	**F**
A1	23	24	2	3	21	0.88	0.91	0.89
A2	22	23	2	3	20	0.87	0.91	0.89
A3	25	22	5	2	20	0.91	0.80	0.85
A4	8	8	1	1	7	0.88	0.88	0.88
B1	6	4	3	1	3	0.75	0.50	0.60
B2	7	8	1	2	6	0.75	0.86	0.80
B3	11	8	4	1	7	0.88	0.64	0.74
B4	4	8	0	4	4	0.50	1.00	0.67
Overall	**106**	**105**	**18**	**17**	**88**	**0.84**	**0.83**	**0.83**

Table 5.4: Accuracy assessment for ITC after refinement using distance-based algorithm in FRIM site according to statistics parameters

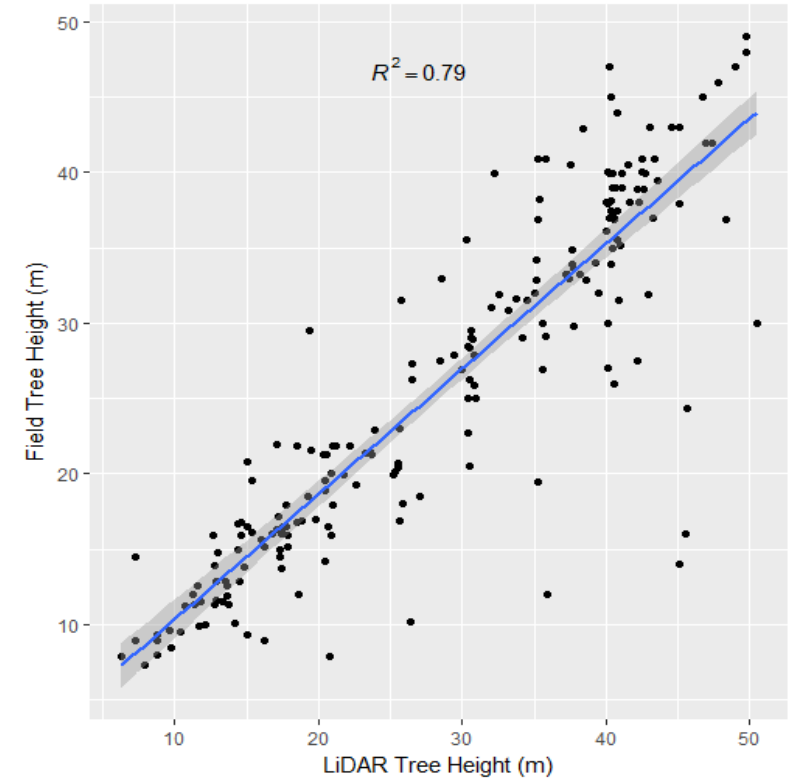
	Number of Trees Detected							
Subplots	**LiDAR**	**Reference**	**FP**	**FN**	**TP**	**r**	**p**	**F**
A1	17	13	4	0	13	1.00	0.76	0.87
A2	21	17	6	2	15	0.88	0.71	0.79
A3	17	15	3	1	14	0.93	0.82	0.88
A4	23	18	6	1	17	0.94	0.74	0.83
B1	12	11	2	1	10	0.91	0.83	0.87
B2	14	12	3	1	11	0.92	0.79	0.85
B3	6	9	1	4	5	0.56	0.83	0.67
B4	8	9	1	2	7	0.78	0.88	0.82
Overall	**118**	**104**	**26**	**12**	**92**	**0.88**	**0.78**	**0.83**

5.4.3 Validation and uncertainty analysis

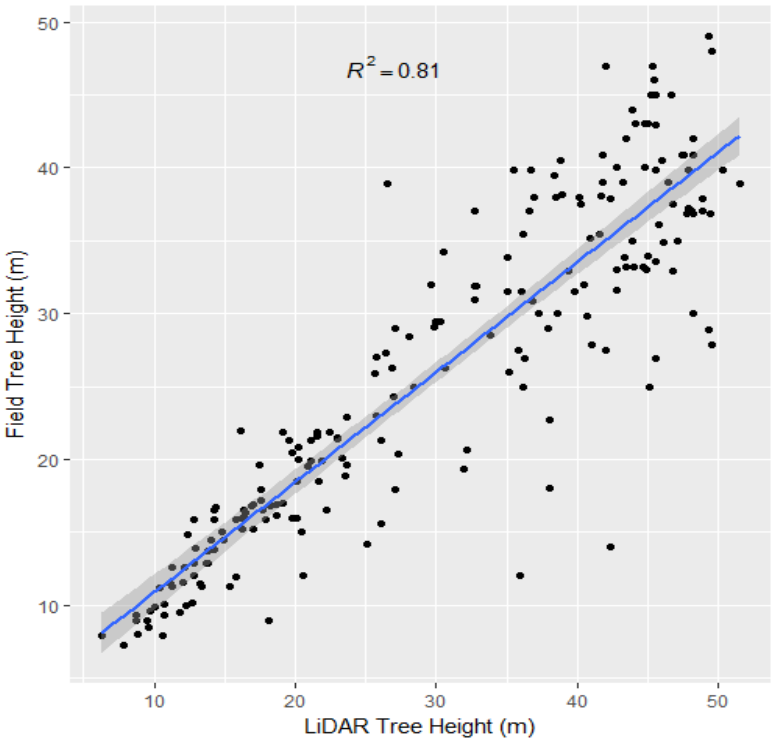
For validation purpose, we compare the ITC delineation results with the watershed segmentation output and after refinement results with the distance-based algorithm output. Tree height, crown size and AGB results were compared. The accuracy of the LiDAR based models is evaluated using the coefficient of variation (R^2). An AGB model developed in Wan Mohd Jaafar et al. (2017) for the study site was used to calculate the LiDAR AGB and the results were compared. One of the advantage from this study is, the results from the ITC delineation can be directly compared to field data such as tree height or crown size since the measurement were taken in 2014, 2 years ahead than the LiDAR data acquisition in 2012. Since the study areas is a primary forest and almost pristine, there are no significant changes in tree size over this short period. It is feasible to locate and compare the crown size and height with the LiDAR data. The AGB value was calculated by using the LiDAR-AGB model as developed by Wan-Mohd-Jaafar et al. 2017 (section 4.5.1) expressed as:

$$AGB_{LiDAR} = EXP[0.57 + 0.61 * \ln CW + 1.55 \ln h80] \, . \quad (5.7)$$

Where CW is the crown width and h80 is the LiDAR 80th percentile derived from the LiDAR point cloud metrics. The relationship between the reference and LiDAR-derived tree detection is shown in Figure 5.14, 5.15 and 5.16. There is a stronger relationship for the delineation results after refinement with the density-based algorithm compare to watershed as the initial method (from 0.79 to 0.81). The correlation between LiDAR-derived height and field reference (Figure 5.14b) is slightly improved after refinement compare to the watershed result (Figure 5.14a). The trend follows in the result for the crown width and AGB estimates between LiDAR and field reference for both method. The pearson-R^2 significantly improved after refinement from both measurement with CW R^2 improved from 0.84 to 0.86 (Figure 5.15a and 5.15b) and AGB estimates R^2 from 0.64 to 0.69 (Figure 5.16a and 5.16b).

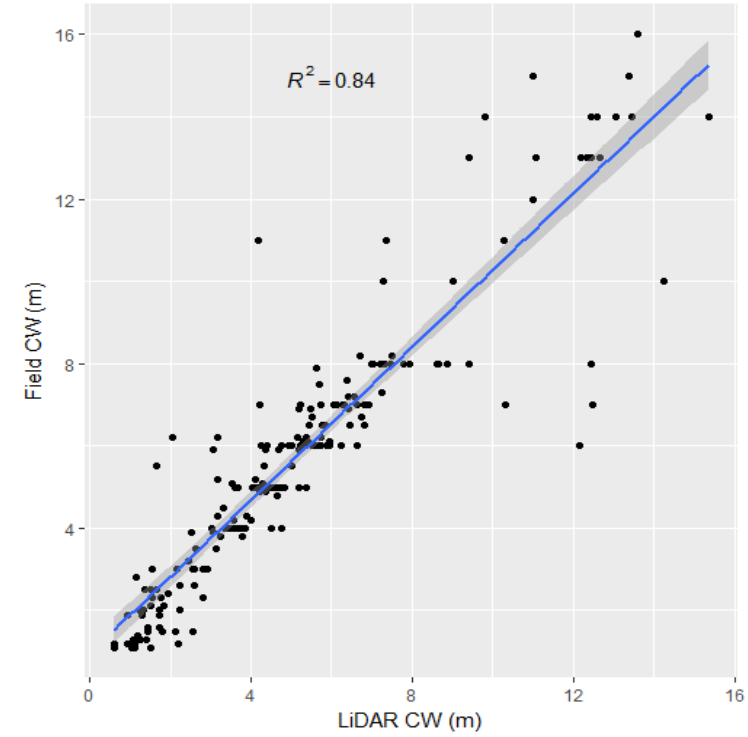


(a)

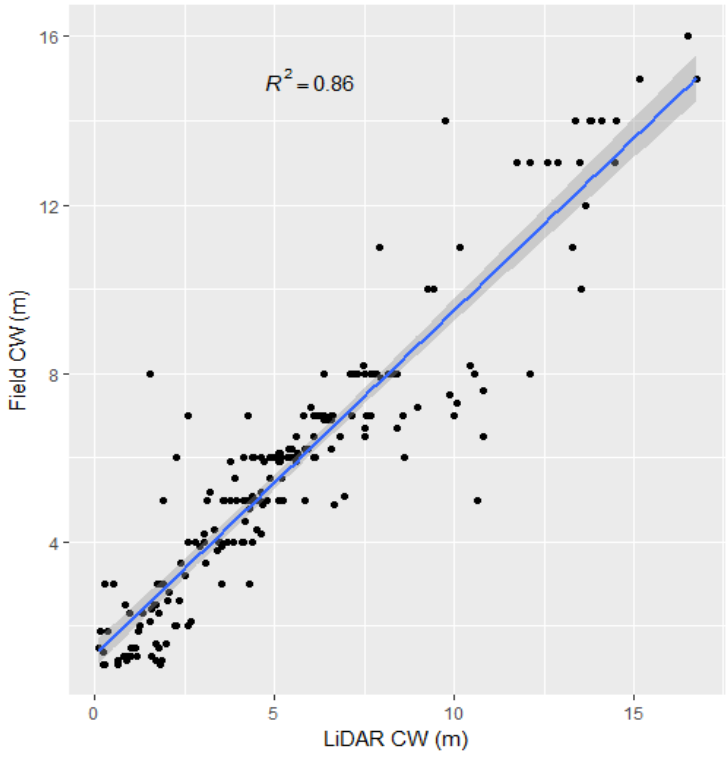


(b)

Figure 5.14. LiDAR-derived tree height versus field reference tree height measured in the field (a) morphological watershed segmentation results and (b) results after refinement with distance-based algorithm

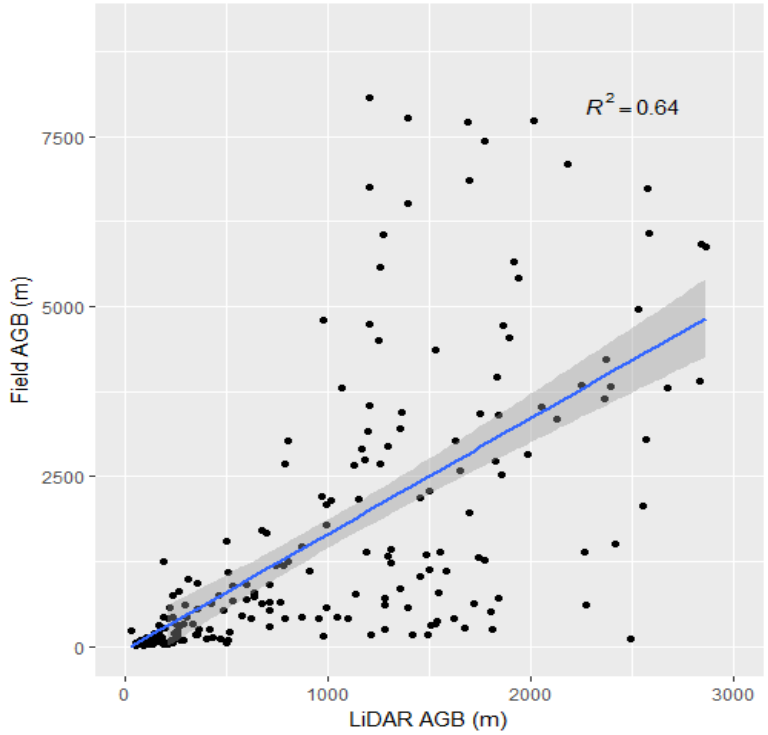


(a)

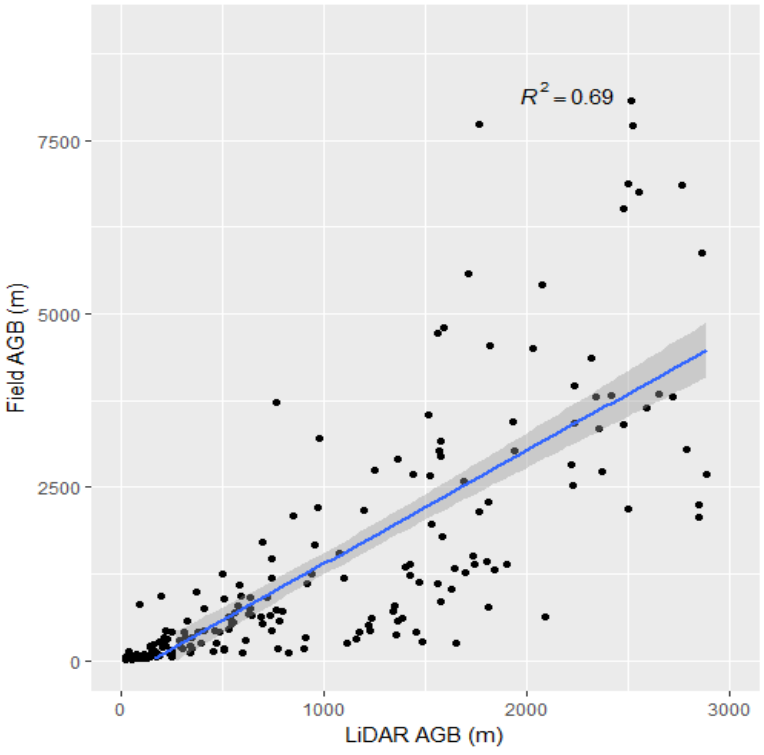


(b)

Figure 5.15. LiDAR-derived CW versus field reference CW measured in the field (a) morphological watershed segmentation results and (b) results after refinement with distance-based algorithm



(a)



(b)

Figure 5.16. LiDAR-derived AGB versus field AGB measured using allometry equation (a) morphological watershed segmentation results and (b) results after refinement with distance-based algorithm

To emphasize the importance of a correctly delineated crown delineation, Popescu et al. (2003) have shown the importance of correctly estimating the crown size and how much the results can improve the estimates of other forest biophysical parameters such as AGB. The average LiDAR-derived crown diameter explained 78% of the variance associated with biomass. The decrease in R^2 values when predicting biomass without using LiDAR-measured crown diameter variables was on average 0.11 with a maximum of 0.24 for regressing biomass where RMSE decreased by up to 7 Mg/ha. In this study, a simple straight-forward assessment was made on how much the value of AGB varies between single delineated trees (from DBA results) and an overlap crown that misclassified as one tree (from watershed segmentation result) (Figure 5.12). The AGB values from DBA detected as two single trees is 1802.2 Kg/tree and 2100 Kg/tree and AGB values from watershed that detected as one tree value is 2458.542 Kg/tree. There's a huge difference in terms of biomass values at tree level with difference of 1443.7 Kg/tree.

5.5 Discussion and Conclusions

In this study, the framework was designed in an efficient way where the detailed examination on the 3D LiDAR points was only focus only for those that need further evaluation based on prior knowledge and not needed for all segments. It was also effective in terms of delineating ITCs in the study plots with good accuracies assessed by both the visual assessment and the quantitative assessment (Table 5.3 and 5.4). The results obtained for individual tree detection differ significantly from study to study. It is difficult to compare the results in this study with those reported in existing literature using different methods since the scenarios in experimental such as LiDAR configuration and the forest morphology are not the same. It is also unclear how much of the variation is caused by the applied methods and forest conditions. A comparative analysis was made between this study approach and CHM-based method as well as point cloud-based methods. It is well reported that CHM-based method produces lower detection accuracy than point based methods because it works on a raster surface model.

Jakubowski et al. (2013), perform ITC delineation by using Object-Based Image Analysis (OBIA) to segment the CHM and point cloud segmentation algorithm by Li et al. (2012) same as the author used in this study, to segment the LiDAR data into individual trees. It is reported as a separate result, with OBIA over-segmented many large trees, but in some cases, it correctly delineated smaller trees that the point cloud approach has missed and the point cloud approach produced larger polygons, especially for the tall and large trees but in overall both methods detected trees well when compared to the ground reference with high Pearson's squared correlation (0.92 and 0.93 respectively). The difference with this study is, the point cloud approach was applied to the 3D LiDAR points from the results of clipping with the polygon segmented from the watershed segmentation. It works to refine the results from the raster-based approach and the results are comparable to the results obtained in other studies using both point cloud and raster-based approaches. With one reported as the accuracies detected for deciduous forest is relatively low compared to conifer forest (Jing et al. 2012; Hu et al. 2014), this can rather be expected the results for tropical forest is very much lower than any other type of forest due to the difference in forest structure, study sites, data used and evaluation method.

In terms of methods used in this framework, in this study the author introduces one of the occasionally used method of assessing the features of the dominant objects in a scene of interest by using semi-variogram and morphological methods together injected to the watershed segmentation algorithm, where most of the time both the semi-variogram and morphological methods have been used separately (Kwak et al. 2007; Zhao et al. 2012; Liu et al. 2016). Three and four scale levels of crown sizes were determined from CHM images of PFR and FRIM site, respectively, and they were consistent with visual observations.

Geometric and structural knowledge of tree crowns was exploited and used to evaluate the initial segments generated by the marker controlled watershed segmentation method. The prior knowledge of tree crowns was represented by a small set of rules. Segments that did not satisfy these rules were identified for further refinement. The identified problematic segments that need to be segmented again contained tree

clusters or tree branches indicated that the number of rules applied was adequate. However, if detailed information of crown of interest is available, more advanced methods are required. There are a few studies that conduct a similar approach by running a second evaluation on the identification of the problematic segments from the watershed segmentation by setting up a few rules on the crown segmented for e.g. Reitberger et al. (2009), used an algorithm called RANdom SAmple Consensus (RANSAC) to identify tree stems within those identified groups of trees, then used those stems as a starting point for clustering by using normalized cut clustering to partition the points within a clump of trees into individual trees. The detection rate reported was less than 70%. Duncanson et al. (2014) on the other hand refine the first watershed delineation by further split the point cloud vertically, and then ran another set of watershed delineation over the lower layers by using trough algorithm to generate the multi-layered crown delineation. The approach is not suitable for tropical forest because the boundaries of the layers are fuzzy and difficult to identify (Jaskierniak et al. 2011).

In this study, a distance-based algorithm proposed by Li et al. (2012) was used to improve the segmentation results by making some modification to the spacing threshold rules and a few parameters added based on our knowledge of the forest in our study areas. In this algorithm, the uncertainty in tree segmentations mainly derives from the spacing threshold. In sparse forest, the tree spacing is large and a relatively larger threshold can be used to isolate the trees but it is difficult to determine an appropriate threshold in dense forest because a higher threshold can result to under-segmentation and a lower threshold can result to over-segmentation. This under and over-segmentation problem can be reduced by using an adaptive threshold by assuming that taller trees have larger crown diameters and hence larger spacing at the upper level. The author also incorporates more classification rules (e.g distance, distributions of points and spatial structures) in the morphological opening procedures and semi-variogram statistics to find the range of crown sizes as discussed in section 5.3.1.

Sufficient LiDAR points is required for the detection of the number of trees to generate the tree profiles. High LiDAR point density is important for accurate generation of the tree profiles. If the LiDAR point density is relatively low (e.g. 1 point/m^2) refinement on the method is needed to ensure that the proposed framework could still yield useful delineations. For future work, it would be interesting to determine the minimal LiDAR point densities that could be used to produce satisfied ITCs delineation results in order to acquire LiDAR data most efficiently for operational consideration.

In conclusion, the proposed methods successfully yielded crown maps having good consistency with manual and visual interpretation. With the advantage of LiDAR data to characterize the tree crown profiles, the accuracy and efficiency of ITC delineation are improved using the proposed five-step framework.

CHAPTER 6

Parametric and Non-parametric Approaches for Estimating Individual Tree Attributes in Tropical Forest Using LiDAR Data

A portion of the research in this chapter is going on peer review in the following journal paper:

Wan-Mohd-Jaafar W.S., Woodhouse I.H., Omar H., Breidenbach J, Hudak A.T, Silva C.A., (2017). Parametric and Non-parametric Approaches for Estimating Individual Tree Attributes in Tropical Forest Using LiDAR Data.

Author Contributions:

My contributions in this research include collected, processed and analysed LiDAR data, interpreted results, prepared the manuscript, and coordinated revisions of the manuscript. Iain Woodhouse supervising the work, assisted with interpretation of the results and reviewed the manuscript. Hamdan Omar, assisted in designing the fieldwork experiment, assisted in data collection, and provide access to compilation of primary and secondary dataset and reviewed the manuscript. Johannes Breidenbach, Andrew Hudak and Carlos Silva reviewed the manuscript.

6.1 Introduction

Forest inventory is fundamental to forest planning and forest inventories, ultimately leading to the development of a forest management plans that allow forest managers to provide information about the quality and quantity of forest resources. Often, times these data must cover large areas of land, up to thousands of hectares, so that finding the balance of the amount (and quality) of data required and the cost to collect them can be very difficult. In recent years, the need for cost-effective, accurate forest inventory data has led to new ways of estimating and imputing plot data, as done by the United States Department of Agriculture (USDA) Forest Inventory and Analysis Program (FIA), and this has raised an interest among forestry agency at other countries such as FRIM (Forest Research Institute Malaysia). Remote sensing technologies such as high-resolution satellite imagery, microwave radars and laser scanning allow the surveying of large areas at reasonable costs (McRoberts et al. 2010). Airborne LiDAR (Light Detection and Ranging), can act as a measurement source independent of the variability of field operators, has demonstrated an ability to improve the accuracy of forest inventory parameters such as height and volumes (Maltamo et al. 2009; Tonolli 2011). LiDAR-supported forest inventory has become an important and active topic in forest research even though wall-to-wall LiDAR information across large areas is not regularly available due to high costs and data volumes (Wulder et al. 2012). Within this area of research, comparisons of forest inventories methods using the area based approach, with the individual tree crown (ITC) approach, remain an active research issue (Hyyppä & Inkinen 1999; Breidenbach & Astrup 2014).

Forest inventory information acquired from field surveys and remotely sensed data requires statistical modelling to obtain overall information from discretely sampled observations. Recent studies have emphasized the high potential and feasibility of LiDAR data for performing forest inventories which applications of LiDAR in forestry include: forest attribute and parameter estimation [e.g. AGB, aboveground carbon stock (AGC) and Leaf Area Index (LAI)], 3D canopy structure measurements (e.g. crown diameter, tree volume and tree height), and tree species mapping (Hyyppä et al.

2008; Hudak et al. 2008; Gleason & Im 2012). Various methods have been implemented for using LiDAR data in parametric and non-parametric approaches: ordinary least square (OLS) regression (Gobakken et al. 2012), most similar neighbour (MSN) imputation (Kankare et al. 2013; Gagliasso et al. 2014), gradient nearest neighbour (GNN) imputation (Hudak et al. 2008; Gagliasso et al. 2014; Hudak et al. 2014; Temesgen & Ver Hoef 2015) and Random Forest (RF) (Falkowski et al. 2010; Hudak et al. 2008; Temesgen and Ver Hoef 2015). All of these methods are based on an area based approach.

Of all the methods, OLS regression with stepwise variable selection has been most frequently used for building models between field measurements and LiDAR metrics (Garcia-Gutierrez et al. 2014). OLS regression is a common modelling method in forest biomass and carbon stock estimation using LiDAR data (Li et al. 2014). This method minimizes the sum of squared vertical distance between the observed responses and the responses predicted by linear approximation. However, the fitting and applicability of OLS regression models relies on a number of basic assumptions in relations to the residual distribution which are: independence, normality and constant variance (Montgomery et al. 2012). These assumptions are barely taken into account in most studies especially when dealing with the data that are collected from complex field survey designs (García et al. 2010). Ignoring the model assumptions when fitting OLS regression models, might lead to spatially correlated errors and consequently invalid significant tests (Fox et al. 2001).

Non-parametric approaches such as RF (Hudak et al. 2008) and *k* nearest neighbours (kNN) imputation such as GNN and MSN (Falkowski et al. 2010) are also considered as an alternative to OLS regression, since they do not rely on any distributional assumptions of the data (Packalén & Maltamo 2007). MSN methods have been used for forest aboveground biomass and stem density (Kankare et al. 2013), tree volume estimation, stand density and quadratic mean diameter estimation (LeMay et al. 2008), and cavity tree abundance (Temesgen et al. 2008). GNN methods on the other hand

have been used to estimate important forest structure attributes such as basal area and canopy cover (Pierce et al. 2009) and tree species composition (Ohmann & Gregory 2002) and similarly, RF method have been used increasingly (Hudak et al. 2008; Eskelson et al. 2009; Latifi & Koch 2012; Hudak et al. 2014). Thus, RF and kNN are highly relevant alternatives to deal with non-linear and possibly diverse relationships between independent and dependent variables. kNN allows for both univariate and multivariate predictions of continuous and categorical variables. In forest inventory applications, kNN approaches have been frequently applied in model-dependent frameworks with good results and have also been used to map various forest attributes (Chirici et al. 2012). Several studies have compared the performance of kNN with OLS regression models in temperate and boreal forests (Penner et al. 2013; Gagliasso et al. 2014; Bollandsås et al. 2013). The results varied according to response variable, prediction types and performance measure.

Only a few studies have addressed the use of ITC for forest inventory. One of the approaches was introduced by Peuhkurinen et al. (2007), who derive pre-harvest information on stands to compare ITC with other approaches. Parametric methods were used in their studies to estimate tree properties without differentiation of tree species and conclude that ITC outperformed the area-based approach and sample plot inventories despite the biased results from the underestimation of timber volume. Flewelling (2009) used a probability models for tree count and other parametric methods for estimation of diameter and height on a segment basis by adding trees from the individual stratum observations to aggregated results to overcome the observed underestimation. Lindberg et al. 2008, applied seemingly unrelated regression (SUR) methods to derive tree characteristics on an individual segment basis. They combined the ITC results with the area-based approach by using Euclidean distance based on explanatory variables by adding trees from the nearest neighbouring plot. Breidenbach et al. (2010) on the other hand used an extension of the methods described by Lindberg et al. (2008) and Flewelling (2009); they proposed a new statistical method to handle errors in the ITC algorithms. Their method allows an unbiased prediction of the

response without the additional step of using data from an area-based approach and this approach was called semi-ITC.

Of particular interest is application and validation of such techniques in tropical forest, where the applicability of statistical methods commonly used in LiDAR-based forest inventories are still limited compared with temperate and boreal forests. Given the growing potential of the use of national forest inventory data, and the additional LiDAR information for supporting REDD+ activities in tropical forests (Tomppo et al. 2014; Leitold et al. 2015), it is important to explore modelling methods that fully utilize the attributes of design as fundamental steps towards reliable and accurate estimation forest attributes parameters using LiDAR.

In this study, the author assesses the performance of parametric (linear modelling) and non-parametric methods (kNN and RF) for modelling AGB and three forest attributes, tree height (Ht), basal area (BA) and stem volume (V) by using LiDAR data and ground measurement across the site on Pasoh Forest Reserve and FRIM forest reserve. kNN methods have been used extensively in studies that employ area based approach, there seems to be only one study that applied MSN to estimate attributes of selected single trees (Maltamo et al. 2009) and only a few studies that apply RF in an ITC (Breidenbach et al. 2010; Silva et al. 2016). To the very best of our knowledge, this is the first study that applies parametric and non-parametric approaches on tropical rainforest biomes for modelling AGB and forest attributes parameters in an ITC or segment level context. The ITC delineation results from Chapter 5 was adopted to derive the individual tree LiDAR-derived metrics as the predictor variables to estimate the particular attributes of interest. The statistical performances of the modelling methods were assessed in terms of precision, accuracy and efficiency.

6.2 Materials and Methods

Details of the study and area, LiDAR data acquisition and field ground measurements was explained in detail in Chapter 3 and 4 (Wan-Mohd-Jaafar et al. 2017). Four forest attributes were calculated for each single tree and used as a response variables for the statistical models described as: BA (m^2), V (m^3), Ht (m) and AGB (Kg). Statistics of these attributes is provided in Table 6.1.

Table 6.1: Summary statistics from ground measurement of forest attributes modeled as response variables

Response	Minimum	Maximum	Median	Mean	SD*
BA (m^2)	0.01	0.53	0.06	0.12	0.14
V(m^3)	0.02	7.99	0.65	1.56	1.96
Ht (m)	7.30	49.00	25.00	25.76	11.12
AGB (Kg)	8.70	8067.16	706.05	1666.69	2026.06

*Standard deviation

6.2.1 Modelling methods

Modelling methods used in this study fall into two categories:

1. Non-parametric methods: this method does not depend on theoretical probability distribution, which can predict values that are within range of training data. The k-nearest neighbour (kNN) imputation methods used in this study are MSN, GNN and RF.
2. Parametric methods: Depend on theoretical probability distribution and are explicit, which can predict values that are outside of range of training data. The linear modelling method used in this study is OLS.

In non-parametric approach, the kNN methods work by direct substitution (imputation) of measured values from sample locations (references) for location for which the user desires a prediction (targets). Observations of both response and explanatory variables is labelled as the reference set and observations for which only the explanatory variables are available is termed as the target set. The similarity between the target observation (*i*th) and the reference observation (*j*th) was quantified by means of the Euclidean distances calculated in feature space as:

$$d_{ij} = \sqrt{(x_{i-}x_j)'\,(x_{i-}x_j)}. \quad (6.1)$$

Where x_i and x_j are the feature vectors. As the d_{ij} distances decrease, the similarity between the target and reference observations will increase, and consequently the nearest neighbour of the *i*th target observation is the reference observation located at the shortest Euclidean distance in the feature space.

Weighted sum of the responses ($\hat{y}_i$) taken from the nearest *k* reference observations as follows:

$$\hat{y}_i = \sum_{j=1}^{k} w_{ij} y_j^i \,. \quad (6.2)$$

Where y_j^i, j = 1,2,…*k* is the set of the response variable observations for the *k* reference set elements that are nearest to the *i*th target set elements in the feature space. The *k*-weights associated with the response in Eq. 6.2 were obtained as;

$$w_{ij} = d_{ij}\left[\sum_{j=1}^{k} d_{ij}\right]^{-1} . \quad (6.3)$$

Selection of *k* has an influence on the accuracy of the imputation. In major case, large values of *k* are not recommended since this will shift the predictions towards the sample mean (Hudak et al. 2008). In this study, the author tested the values of *k* ranging from 1 to 10, and the lowest RMSE % value obtained from the cross validation were selected. Meanwhile, in the RF method, since it differs fundamentally from the traditional imputation methods, it requires some elaboration. RF uses classification and regression trees (CARTs) to partition the data into smaller groups with binary splits at each node that are based on single predictor. The classification output from RF represents the statistical mode of many decision trees, hence more robust model

than a single classification tree produced by a single model (Breiman, 2001). kNN analysis were performed using the *yai()* and *impute()* function from the *yaImpute* (Crookston & Finley 2008) R package (R Development Core Team, 2016). While RF was performed using the *randomForest* package (Liaw & Wiener 2002).

Parametric regression is well known and many of its applications in LiDAR-based inventories are given in Naesset (2002) and Hudak et al. (2006). This approach assumes, among others, identical and independently distributed residual, or error terms. By applying the natural logarithmic transformation to response variables (y) and predictor variables (x), the assumption of identically distributed residuals is often met. This log-log transformation was chosen here based on the diagnostic of the variables distribution on the data of both PFR and FRIM site. This was supported on early examination of the data distribution as performed in Chapter 4 (Wan-Mohd-Jaafar et al. 2017). Furthermore, the relationship between forest variables and LiDAR metrics is usually linearized via non-linear transformations using logarithmic transformation of both the regressors and the response variable (Naesset 2002; Li et al. 2008; Naesset & Gobakken 2008). Stepwise, multiple linear regression was used as a method of OLS:

$$\ln \hat{Y} = b_0 + b_1 \ln(X_1) + b_2 \ln(X_2) \ldots . b_z \ln(X_z) . \quad (6.4)$$

Where $\hat{Y}$ is the predicted value of a dependent variable and $X_1 \ldots X_z$ are independent variables (Table 6.3). In the preliminary model, the independent model was fit independently by attribute. These predictors had to be statistically significant ($p \leq 0.05$) and have a variance inflation factor < 5 to reduce multicollinearity as discussed in section 4.5.1. A key assumption under which OLS is a best linear unbiased estimator is that the residuals are independent. An advantages and disadvantages of the selected modelling methods are given in Table 6.2. The author also used the *lm* package (R Core Team 2016) for OLS.

Table 6.2: Advantage and disadvantage of modelling methods used in the study

Method	Advantage	Disadvantage
Nearest Neighbour	• Easy to understand and implement • Less assumption required • Do not rely on any probability distribution • Predictions are always within biologically reasonable bounds	• The optimal number of reference observations (k) and the type of distance metric used in a single imputation need to determine • The reference data need to cover well the distributions of predictor variables (or else the accuracy could be affected)
Random Forest	• Efficient especially when the number of predictor variables is very large and interactions and correlations among variables are complex • Does not require specification of a functional form	• In classification application, RF approach tend to be biased towards the most frequent classes. • Cannot extrapolate beyond the training data and may not interpret well for few samples conditions • Viewed as "black-box" meaning there is no model with which to illustrate the results
Ordinary Least Square	• Most popular and used method • Strong theoretical background • Can extrapolate	• Require many assumptions • The interaction between predictor variables need to be check

6.3 LiDAR Metrics and Variable Selection

The candidate LiDAR metrics was computed using rMetrics function in the rLiDAR package (Silva et al. 2015). Full details of this procedure were explained in section 4.3.3. However, since the selection of LiDAR metrics in this chapter based on the automated individual tree crown delineation procedures, there are additional LiDAR

metrics considered as a predictor variable based on crown structure as listed in Table 6.3. Stepwise variable selection procedure was implemented with the candidate LiDAR metrics to obtain a final set of predictor variables for every modelling method. Since there is no variable selection technique that can be applied commonly to all the selected methods, the stepwise procedure based on OLS regression was used. The stepwise variable selection procedure (stepAIC) was implemented by the function *lm* in R package *MASS* (Venables & Ripley 2002). The preferred models were selected based on Akaike Information Criteria (AIC). Then, a model in which each predictor variable has high significance in consideration of interaction terms and multicollinearity by variance inflation factor (VIF) by the function *vif* in R package faraway (Faraway 2014) were selected. Stepwise algorithm has always been related to its inefficiency when predictor variables are strongly related, which is the case for LiDAR height metrics and density metrics (Harrel 2015). R^2 improvement technique was further evaluated on the selected models by McRoberts et al. 2013. This is the same repeated procedures as done in developing the AGB predictive model in Chapter 4. The final set of predictor variables obtained is given in Table 6.4. The variable selection was repeated for log transformed response variables and predictors (based on the initial test results) (Table 6.5)

Table 6.3: Candidate of Individual tree crown metrics

Variable	Description
Hmax	Maximum Height
Hmin	Minimum Height
Hmean	Mean Height
Hsd	Standard deviation Height
Hvar	Variance Height
Percentile Height (HTH5, HTH10, HTH15, HTH20, HTH25, HTH30, HTH35, HTH40, HTH45, HTH 0, HTH55, HTH60, HTH65, HTH70,	Percentile Heights 5^{TH},…., 99^{TH} of all returns

HTH75, HTH80, HTH85, HTH90, HTH95, HTH99	
CL	Crown Length
CBH	Crown Base Height
CW	Crown Width
CRatio	Crown Ratio
CPA	Crown Projection Area
CSA	Crown Surface Area
CV	Crown Volume

Table 6.4: List of LiDAR metrics according to response variable selected via stepwise variable selection

Response Variable	Adj.R^2	RMSE	Predictors
BA (m^2)	0.39	44%	Hmin, HTH95, CPA
V(m^3)	0.51	38%	HTH90, CPA, CV
Ht (m)	0.86	12%	HTH95, HTH85
AGB (Kg)	0.71	21%	HTH85, CW

Table 6.5: List of LiDAR metrics according to response variable selected via stepwise variable selection with logarithmic transformation for OLS regression

Response Variable	Predictors
BA (m^2)	LN(Hmax), LN(Hmean), LN(CPA)
V(m^3)	LN(Hmean), LN(Hsd), LN(CPA), LN(CV)
Ht (m)	LN(Hmax), LN(Hmean), LN(HTH80)
AGB (Kg)	LN(Hmax), LN(CW)

Since there is no common variable selection technique for every modelling method in this study, this will raise a question as to whether the stepwise variables selection based on the linear modelling method favours linear models over imputation and RF methods. To investigate this issue, variable selection procedure by the function *varSelection* in R package *yaImpute* (Crookston & Finley 2008) was implemented as a non-parametric variable selection technique. This backward feature selection algorithm began with all candidate predictors and deleted them one at a time by computing the mean distance between the observed and the predicted values. By weakening the prediction accuracy, the predictor could be removed from the predictor set. Thus, the predictor related to the largest mean distance was discarded in every iteration round. This variable selection procedure was repeated 100 times, and variables that were selected by the function that had less than a 0.9 correlation, with all the other selected variables used to predict the responses similar to the study conducted by Hudak et al. (2008), which aimed to restrict redundancy. Figure 6.1 shows the best variables according to the mean distance. This variable sorted from the function is almost similar as variables selected from the stepwise selection procedures; for efficiency, the stepwise method was used in this study to select the best variables

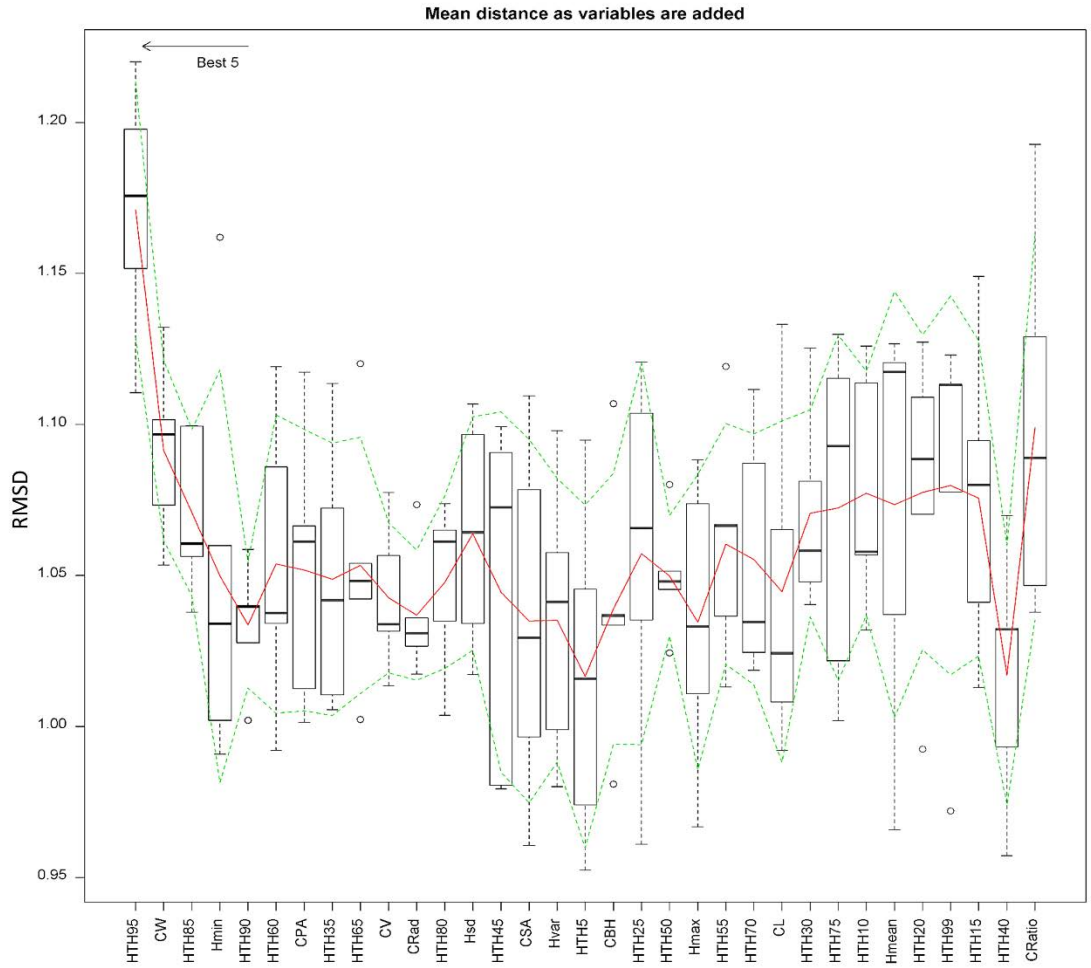


Figure 6.1. Mean distance between observed variables and predicted variables selected from *yaImpute* variable selection

6.4 Model Performance Measure

Simulation by cross-validation with some different numbers of training data (*n*) was used to assess the performance of the prediction methods and the effect on *n* of using the predictor variables via best subset variable selection. The pooled sample dataset from two forest sites (N= 200) was randomly split into 2 groups. One group (80%) was the observed set of plot observations used as the training data (*n*), while the other group (20%) was the validation (target) data (*v*). 10-fold cross validation was also implemented using the predictor variables selected for imputation to examine if the stepwise variable selection favours the linear modelling methods. Performance

measure of the prediction methods by response variables and the two prediction types were computed for each 10-fold cross validation run using bias and RMSE.

1. Bias for point prediction calculated in terms of percentage:

$$Bias\ (\%) = \frac{[\frac{1}{vE}\sum_{e=1}^{E}\sum_{k=1}^{v}(\hat{Y}_{k,e} - Y_{k,e})]}{\acute{Y}} x100 \quad . \tag{6.5}$$

Where E is the number of simulations, v is the number point predictions for the *e*th simulation, $\hat{Y}_{k,e}$ is the prediction at the *k*th location for the eth simulation, $Y_{k,e}$ is the true value at *k*th location for the *e*th simulation, and $\acute{Y}$ is the true average value of a particular response variable.

2. Bias correction after back-transformation Linear model

   The predicted values for each response variables (BA, V, Ht and AGB) obtained from the cross-validation were corrected for bias (due to natural logarithm transformation) because applying inverse natural logarithm transformation to convert the predictions back to the natural scale introduces a negative bias (Hudak et al. 2006). By adding one half of the residual variance to the prediction on the natural logarithmic scale, one can approximate the bias. Multiplying the prediction by EXP (0.5 x MSE) derived the correction factor (CF); MSE is the mean squared error of the residuals (Baskerville, 1972). The MSE were substituted into the equation to calculate correction factors of each predicted response variables and multiplied with the back-transformed predictions:

$$\hat{Y}_{corrected} = CF \text{ x } EXP(\hat{Y}) \ . \tag{6.6}$$

3. Root Mean Square Error (RMSE) measures the difference between true and predicted values. For point prediction, the RMSE is calculated as:

$$RMSE = \sqrt{\frac{1}{v}\sum_{k=1}^{v}(\hat{Y}_{k,e} - Y_{k,e})^2} \quad . \qquad (6.7)$$

6.5 Results

6.5.1 Response variables and Predictor variables relationship

From the individual tree crown delineation results, a total of 200 sample trees with an average height of 23 meter that have been correctly matched with the field data were selected for modelling. BA varied between 0.01 – 0.53 m^2, V for individual trees ranged from 0.02 – 7.99 m^3, Ht ranged from 7.30 – 49.00 m and AGB varied between 8.70 – 8067 Kg/tree (Table 6.1). Table 6.6 shows the correlation coefficients between each response and corresponding selected predictors via the stepwise selection method. The higher the correlation coefficients between response and predictor variables, the better prediction performance. Highly correlated coefficients e.g. Pearson's r>0.9 could introduced multicollinearity.

HTH95 showed high correlation coefficients with Ht but low correlation with BA. HTH85 showed high correlation with Ht and AGB, meanwhile CV showed high correlations with V, with 0.72. In addition to correlation, the scatterplots between responses and the corresponding predictors are in Figure 6.2. From the first screening, since a nonlinear trend shown on the data distribution, natural logarithmic transformation was applied to the response variable and the predictors, the best predictor variables for every response is slightly different according to the correlation (Table 6.7). For BA and V, CPA remain as one of the best predictors. For Ht prediction, there is no common predictor than the first fitting to the stepwise selection, which is HTH95 and HTH85 found as the best predictors based on r^2 valued and for AGB prediction CW remain as one of the best predictors. The scatterplot between responses and the corresponding predictors after log transformation are in Figure 6.3.

Table 6.6: Correlation coefficients between responses (BA, V, Ht and AGB) and predictors (LiDAR metrics) from stepwise selection

		Basal Area	
	BA	Hmin	HTH95
Hmin	0.39		
HTH95	0.31	0.26	
CPA	0.46	0.32	0.39
		Volume	
	V	HTH90	CPA
HTH90	0.51		
CPA	0.64	0.56	
CV	0.72	0.52	0.79
		Height	
	Ht	HTH95	
HTH95	0.86		
HTH85	0.85	0.82	
		AGB	
	AGB	HTH85	
HTH85	0.61		
CW	0.60	0.69	

Table 6.7: Correlation coefficients between responses (BA, V, Ht and AGB) and predictors (LiDAR metrics) from stepwise selection with log-log transformation

		Basal Area		
	LN(BA)	LN(Hmax)	LN(Hmean)	
LN(Hmax)	0.29			
LN(Hmean)	0.31	0.26		
LN(CPA)	0.26	0.12	0.24	
		Volume		
	LN(V)	LN(Hmean)	LN(Hsd)	LN(CPA)
LN(Hmean)	0.67			
LN(Hsd)	0.56	0.64		
LN(CPA)	0.54	0.56	0.53	
LN(CV)	0.52	0.52	0.49	0.59
		Height		
	LN(Ht)	LN(Hmax)	LN(Hmean)	
LN(Hmax)	0.80			
LN(Hmean)	0.78	0.77		
LN(HTH80)	0.79	0.78	0.77	
		AGB		
	LN(AGB)	LN(Hmax)		
LN(Hmax)	0.62			
LN(CW)	0.54	0.60		

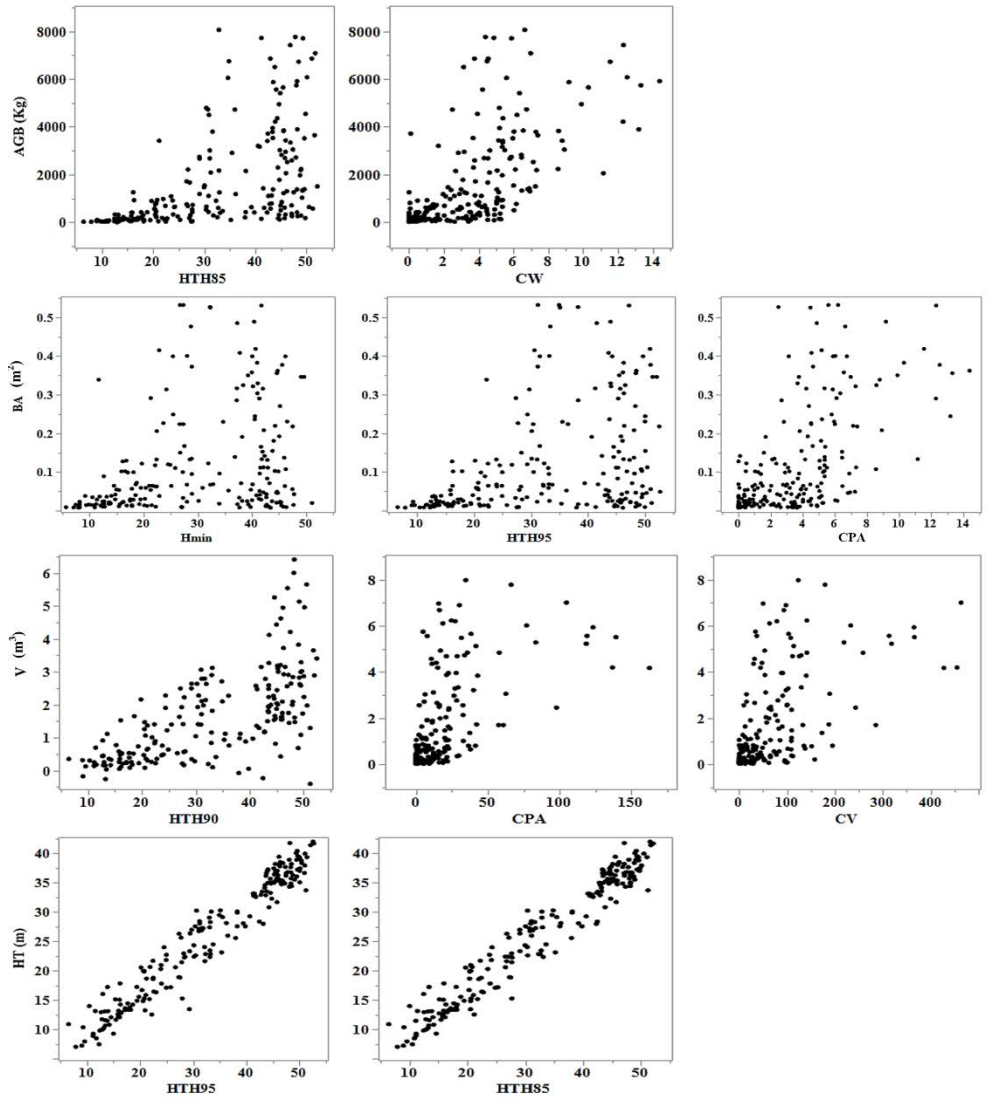


Figure 6.2. **S**catterplots matrix of response variables by predictor variables via stepwise method

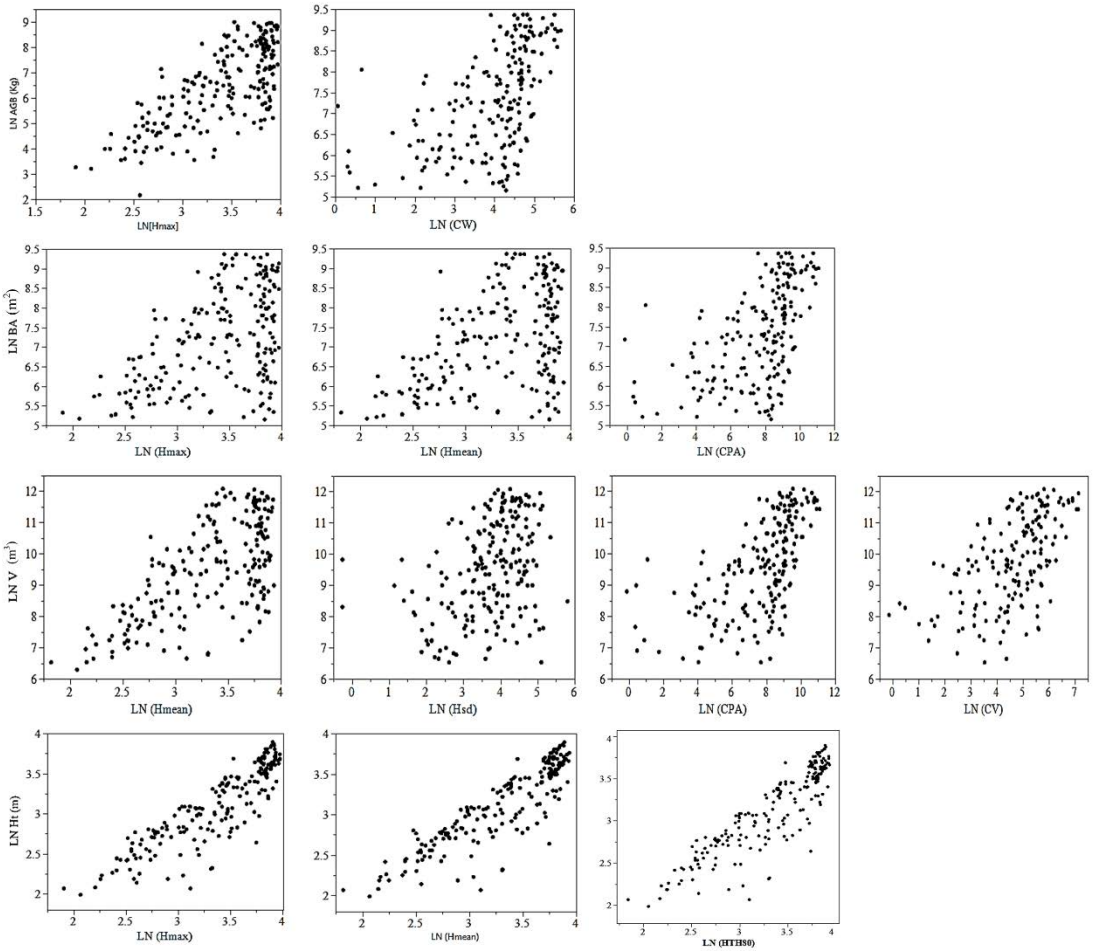


Figure 6.3: Scatterplots matrix of response variables by predictor variables via stepwise method with natural logarithmic transformation.

6.5.2 Performance Measures

Table 6.8 shows the performance measures in 10-fold cross-validation with the predictor variables via stepwise variable selection. From the combination of 4 response variables (BA, V, Ht and AGB) and prediction type, in most cases, OLS showed better performances than the other model types in terms of RMSE and bias and were comparable among each other except for BA prediction. For BA, RF showed the best RMSE in prediction. RF reduced RMSE by 20.8 % over GNN and MSN, and 32.1% over OLS. For V, OLS outperformed the rest of the other models in RMSE. OLS decreased RMSE by 32.7 % over GNN, 31.8 % over MSN and 29% over RF

respectively. For Ht, OLS once again provided the lowest RMSE in model prediction. It reduced RMSE by 56.39 % over GNN, 56.3% over MSN and 51.74% over RF and OLS also produced the best RMSE in AGB, which was improved by 36.2 % over GNN, 36.1% over MSN and 35.8% over RF. RF leading with better performance among the non-parametric approach. GNN and MSN, showed poorer performances with larger RMSEs in most cases. Overestimation was found in most cases, giving positive bias. OLS, GNN and RF underestimated BA and V while MSN underestimated AGB, giving negative bias value. The *p*-values of the *t*-test for testing the unbiasedness of the biases are under the bias values in the parentheses in Table 6.6. Several cases had *p*-values less than 0.001 and 0.05, which means that they are statistically different from zero and significant difference does exist. Several cases accept the null hypotheses ($p > 0.05$), for example all model predictions reported for BA and V but mostly indicated small biases, with low Adj-R^2. But prediction model with biases fit the prediction well with adj-R^2 between 0.79-0.84 for response variable Ht and 0.54-0.60 for AGB.

Table 6.8: Performance measure by response variables and prediction method via stepwise variable selection

Response	**Performance Measure**	**OLS (Back Transformed)**	**GNN**	**MSN**	**RF**
BA (m^2)	Adj-R^2	0.38	0.32	0.34	0.36
	% Bias	-0.20(0.06)	-0.09(0.15)	0.05(0.14)	-0.06(1.32)
	RMSE (m^2)	1.40	1.20	1.20	0.95
V (m^3)	Adj-R^2	0.42	0.38	0.39	0.40
	% Bias	-0.20(0.09)	0.08(0.21)	0.86(0.19)	-1.06(0.18)
	RMSE (m^3)	1.03	1.53	1.51	1.45
Ht (m)	Adj-R^2	0.84	0.79	0.79	0.83
	% Bias	0.15(<.0001)	0.23(<.0001)	0.41(<.0001)	0.07(<.0001)
	RMSE (m)	2.22	5.09	5.08	4.60

AGB (Kg)	Adj-R^2	0.61	0.55	0.54	0.60
	% Bias	0.05(<.005)	6.68(<.005)	-10.42(<.005)	3.42(<.005)
	RMSE (Kg/tree)	96	150.39	150.33	149.56

6.5.3 Prediction Range

Boxplots of prediction and observations as reference are shown in Figure 6.4. The horizontal lines on the quantile boxplot correspond to the quantiles in the distribution output. The black dotted indicated the potential outliers (disconnected points). All parametric model (OLS) and non-parametric model in this study produced positive predictions, which is acceptable and makes sense for forest inventory attributes.

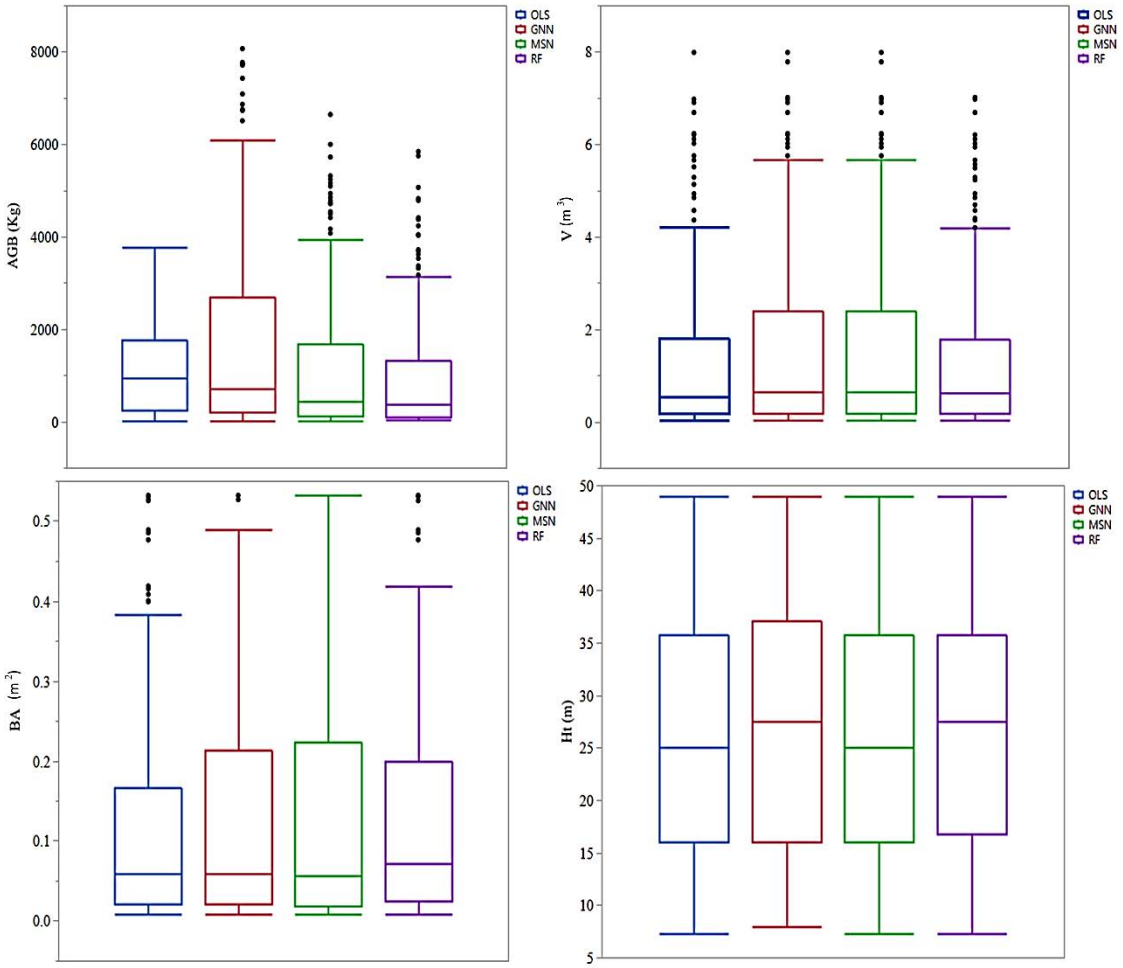


Figure 6.4: Boxplots of predictions by each modelling method tested, for each response variable.

6.5.4 Parametric vs non-parametric

Pearson correlations between observed and predicted individual tree BA, V, Ht and AGB using linear model, NN model and RF-regression based model are shown in Figure 6.5 – 6.8. GNN and MSN performed poor especially on BA, V and AGB prediction. RF gives a better result among the non-parametric approaches. This is probably because, GNN and MSN are performed as imputation method, while RF runs as a classification method. Linear regression OLS outperformed the rest of the methods in terms of adj-R^2 and RMSE. Imputation errors are almost always greater than regression errors because the errors do not result from a least-squares minimization like in regression (Stage & Crookston, 2007). Imputation errors are instead calculated from the pool of observations and include a distance component.

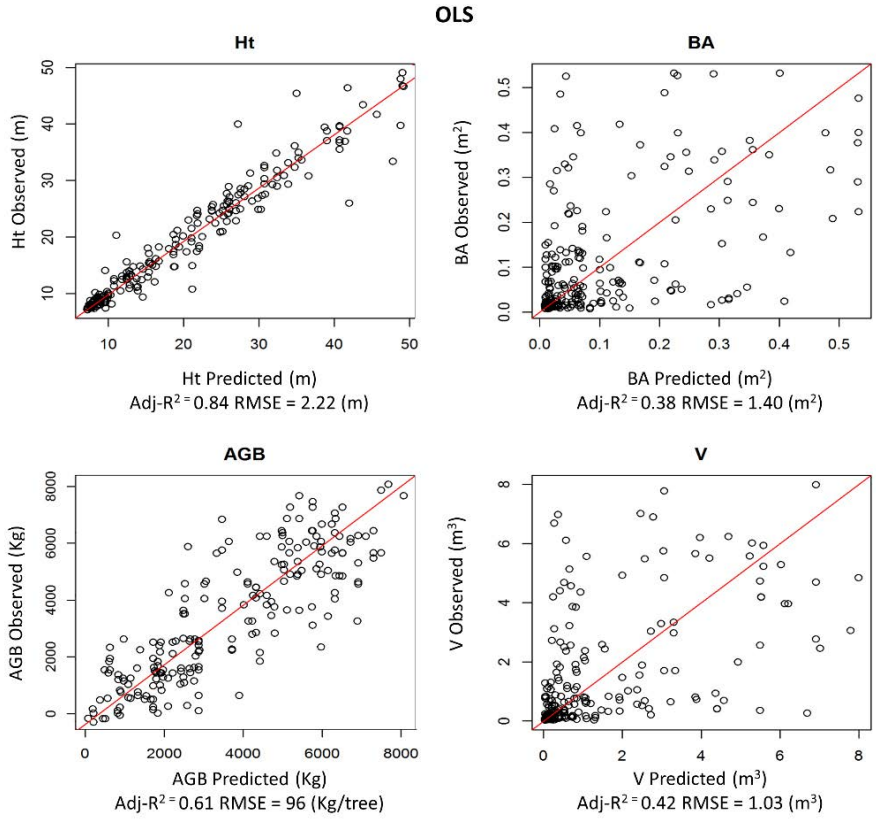


Figure 6.5: Scatterplots of predicted versus observed variables from OLS methods.

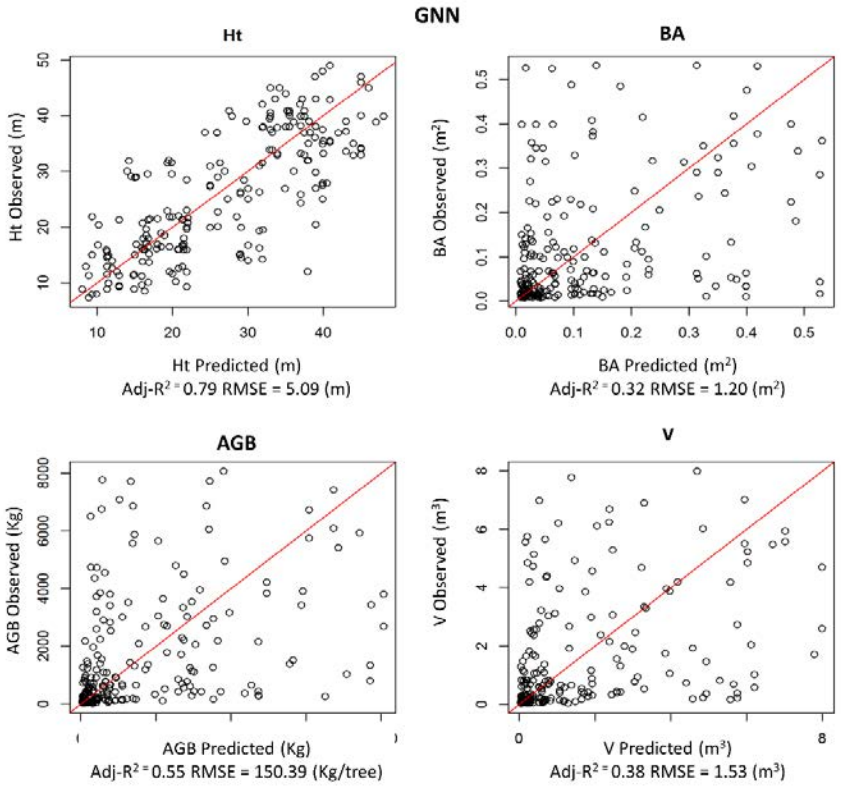


Figure 6.6: Scatterplots of predicted versus observed variables from GNN methods.

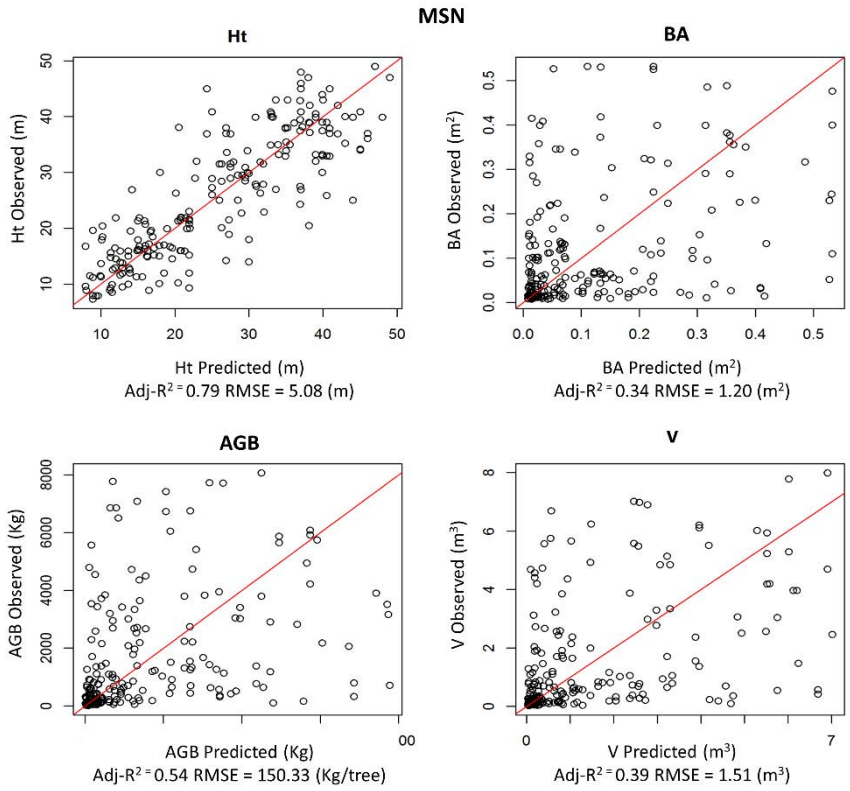


Figure 6.7: Scatterplots of predicted versus observed variables from MSN methods.

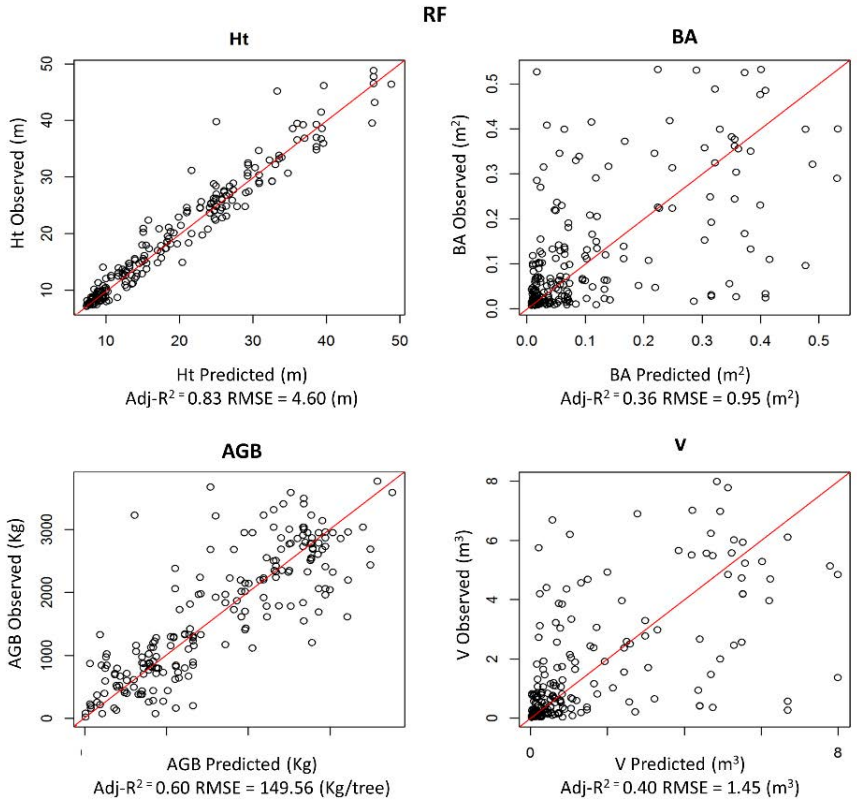


Figure 6.8: Scatterplots of predicted versus observed variables from RF methods.

6.5.5 Bias correction after back-transformation

The bias correction needed to correct the negative bias that was introduced by converting the inverse natural-logarithm back to the natural scale and this was approximated by adding one half of the residual variance to the prediction on the natural logarithm scale. By taking AGB prediction for example, the mean square error (MSE) from a ten-fold cross validation was 0.889, were substituted into the equation to calculate the correction factors (CF) of 1.36; after multiplied with the back-transformed predictions, these CF slightly underestimated the mean AGB by 0.14 Kg/tree (Table 6.9). In general, the distribution of predictions better matched the distribution of observation after bias correction than before bias correction.

Table 6.9. Summary statistics of observed and predicted variables values before and after bias correction for the inverse natural logarithm transformation.

Response	Min	First quartile	Median	Mean	Third quartile	Max
BA						
Observations	0.01	0.02	0.06	0.12	0.17	0.53
Predictions after back-transformation	0.03	0.06	0.11	0.11	0.15	0.48
Predictions corrected for bias transformation	0.01	0.07	0.14	0.12	0.17	0.29
V						
Observations	0.03	0.18	0.65	1.56	2.37	7.99
Predictions after back-transformation	0.05	0.72	1.44	1.59	2.03	7.19
Predictions corrected for bias transformation	0.06	0.62	1.39	1.56	2.00	6.92
Ht						
Observations	7.30	16.00	25.00	25.76	35.50	49.00
Predictions after back-transformation	7.03	15.86	27.15	25.78	35.40	42.01
Predictions corrected for bias transformation	7.78	16.86	25.72	25.77	35.27	40.34
AGB						
Observations	8.70	199.29	706.05	1666.69	2684.35	8067.16
Predictions after back-transformation	23.80	497.40	1725.20	1704.80	2490.20	6079.90
Predictions corrected for bias transformation	19.93	452.30	1654.10	1666.55	2491.90	6344.70

6.6 Discussion

6.6.1 Performance Measure

One modelling methods showed superior performance for all cases investigated in several previous studies (Shin et al. 2016; Hudak et al. 2014). OLS was always the best method in RMSE for forest data in Li et al. (2014). RF was the best for basal area in root mean square distance (RMSD) in Hudak et al. (2008) and was the best for 9 forest attributes in RMSD in Hudak et al. (2014). Eskelson et al. (2009) indicated that RF imputation showed the best results in terms of RMSE for mean annual change in basal area, stems per ha, volume, and biomass and those current forest attributes using climate, topography, and satellite data. These findings supported the result in this study where RF was the best for BA prediction in terms of RMSE. However, the modelling method varied by response variable, performance measure, and prediction type. Pierce et al. (2009) indicated that the best modelling method among GNN and OLS for vegetation and fuel variables varied by response variables and regions where the models were applied. Corona et al. (2014) concluded that correlation between response variables and auxiliary variables affects the performance of a modelling method.

From several previous studies, all models reported biased at least one predicted response. Ver Hoef & Temesgen (2013) claimed that, in the resampling of real-forest data, there appeared to be some bias for total prediction by kNN methods with Mahalanobis distance (k = 1 or k = 5) except with spatially balanced sampling. In Eskelson et al. (2009), RF was biased for prediction of mean annual change in basal area, volume, and biomass by using climate, topography, and satellite data. As stated earlier, RF consistently produced larger biases than other kNN methods (Breidenbach et al. 2010). In this study, kNN methods with k = 5 had lower RMSE. Similar results have been reported by several other studies. Muinonen et al. (2001) reported that MSN had lower RMSE% than GNN whereas k increased to 5 for plot-level tree volume estimation.

In Temesgen & Ver Hoef (2014), MSN with *k*=5 had the smallest RMSE for biomass and basal area estimation, followed by MSN with *k* = 3 and MSN in Gagliasso et al.(2014). In terms of bias, however, kNN methods with larger *k* do not always provide less bias than kNN methods with smaller *k*. As with our study, Temesgen & Ver Hoef (2014) and Ver Hoef & Temesgen (2013) reported that the effect of *k* on bias of kNN varied by response variable or dataset used. MSN with *k* = 3 showed the lowest bias for biomass and basal area estimation compared to MSN with *k* = 5 (Gagliasso et al. 2014).

Predictions provided by linear and kNN methods were within the biologically reasonable bounds, not giving negative values or unrealistically high values of forest attributes examined. Linear modelling OLS showed superiority in most of forest parameters prediction. Transformations of dependent variable and independent variables by natural logarithmic could be one of the reason that results to the best performance overall because log transformation can decrease the variability of data and make data more closely to the normal distribution, this is supported by Shin et al. 2016 where transformations are recommended for the OLS method to offer some improvement to the model and other modelling methods for forest attributes prediction.

6.6.2 Influence of variable selection technique on model performance

Other than the variable selection method proposed by Crookston & Finley (2008) and used in this study for imputation, there are several approaches to select predictor variables for the NN imputation method. An algorithm to minimize relative RMSE for NN imputation, using transformation and stepwise optimization, was proposed by Maltamo et al. (2006). Hudak et al. (2008) repeated RF to discard the least important predictor variable, which is similar to backward stepwise variable selection in a multiple regression model. Latifi et al. (2010) claimed that the prediction performance depended on the combinations of NN imputation methods and variable selection procedure. With NN imputation using distance metrics such as Euclidean and Mahalanobis, genetic-algorithm (GA)-based variable selection produced better

prediction performance. By contrast, full dataset, i.e., without variable selection, outperformed GA-based variable selection when using MSN and RF imputation. The backward stepwise variable selection gave the lowest performance compared to GA-based variable selection and full dataset. GarciaGutierrez et al. (2014) compared 3 variables selection techniques (stepwise, best subset, and GA selections) for estimating some forest stand variables using LiDAR. GA was reported to perform better than other techniques, based on BIC of 3 regression models. Packalén et al. (2012) concluded that variable selection is an essential part of NN imputation. An algorithm using optimization to minimize relative RMSE was the best variable selection strategy compared with the full dataset, another using canonical correlation analysis, and another using RF importance. But the full dataset surpassed the algorithm with canonical correlation analysis and that using RF importance. The results comparing the 2 variables selection techniques (stepwise and varSelection) in the present study showed almost balanced variables and for time-efficiency, this study fused stepwise variable selection for all models. However, our results cannot be generalized compared with the previous study because the dataset differ, and the combination of modelling methods and variable selection techniques in our study were different from those of the others. It would be more meaningful and practical to determine which combination of variable selection technique, and size of *n,* provides the best performance for NN imputation. From this point of view, our comparison is not enough to determine the optimal combination because the comparison was implemented using only 1 variables selection techniques and 1 training dataset, size of 200. This should be further studied in the future.

6.7 Conclusion

This study suggests that depending on what forest managers want to study and find out about their forest, various methods can be used. Number of statistics that can be computed from the LiDAR metrics is almost limitless. The selection of prediction variables in parametric method is generally using empirical approaches such as through stepwise process, best subsets or data-driven for model development. If a forest manager would like to know just one piece of information, the advantages of

parametric method for model-fitting with limited data is seen relevant and could be used in a cost-effective way. However, if multiple forest parameters are desired, non-parametric method is an efficient approach. Because non-parametric approach such as RF does not require the specification of a model and takes full of advantage of a dataset including highly correlated predictors. The implementation of RF for example, on a large area does not exposed to computing challenges in terms of expertise or processing time. RF predictions had precision and efficiency comparable with parametric- OLS method and it appears to be a promising alternative to parametric method for the operational prediction of forest attributes from LiDAR. These results could be beneficial to FRIM in future forest plans to predict biomass and other forest attributes at a project level scale on any FRIM maintained forest.

The results from this study also helps to justify the ITC delineation results achieved in Chapter 5 and the AGB predictive model at tree level derived in Chapter 4 (Wan-Mohd-Jaafar et al. 2017) in terms of LiDAR metrics as predictor variables which interact with the response variables shown comparable to the variables derived from modelling in this study. This study has examined the performance of the statistical methods to estimate different response variables in a tropical forest environment. It is recommended that this study be repeated in other forest environments, or include multitemporal LiDAR, and/or hyperspectral imagery, to analyse further at species level. Comparison of ITC to an area-based approach is also recommended.

CHAPTER 7

Conclusions

7.1 Summary

Airborne LiDAR data has become increasingly accessible over the past decade. As the technology continues to develop, it will provide valuable information for managers and decision makers worldwide for a variety of forestry applications (Merrick et al., 2013). This thesis has investigated the novel techniques of airborne LiDAR-based assessment for individual tree characterization and analysis in a tropical forest region. The major conclusion of this research is that the author provides evaluated techniques for aboveground biomass (AGB) quantification as the main parameter of interest to derived from LiDAR data using approaches that segment individual tree crowns. The scope of the methods arranged from simple regression between LiDAR-derived height metrics and AGB at tree level based on trees that were mapped in the field to methods of automated tree crown delineation and finally modelling by using machine learning approaches for AGB and forest attributes such as Basal Area (BA), Volume (V) and Height (Ht) estimation. The success of this research is mainly attributed to the novelty of LiDAR as the main tool, which has been proved to be effective to estimate important forest parameters. This thesis is of scientific value for understanding the benefits provided by high

density LiDAR data (7.5 – 8.8 pts/m^2) in individual tree analysis. The implementation of the proposed methods to broader forests is feasible. The novelty and major contributions of this research have fulfilled the three aims stated in Section 1.3 which can be summarised as follows:

1) Analytic flows for exploring the relationship between LiDAR metrics and field tree attributes at the individual tree level was developed for a tropical forest site. Due to the structural complexity of tropical forests, a thorough identification of the relationship between AGB measured in the field and LiDAR derived metrics is important to increase the accuracy of predictive models as an initial step before progressing to a more complex analysis.
2) A Framework to improve accuracy and efficiency of Individual Tree Crown (ITC) delineation from LiDAR data was developed; (1) multiscale smoothing on a canopy height model (CHM) was proposed to determine the best scale for identifying individual tree crowns and, (2) point cloud method was applied to refine the individual tree crown segments.
3) The performance of parametric and non-parametric approach for modelling individual tree AGB, BA, V and Ht using LiDAR data over tropical forest site was compared and assessed.

Detailed findings of this study are described as follows:

- Chapter 4 (Wan-Mohd-Jaafar et al. 2017, *Journal of Tropical Forest Science*) 29(4) p 465-484: LiDAR height percentile h80 and crown width measurement, CW was found to best fit the data as evidenced by an Adj-R^2 value of 0.63, root mean square error (RMSE) of the model of 14.8%, and analysis of the

residuals. The AGB model was developed based on estimated AGB at the field site and LiDAR data using the following methodology: (1) pooling of both field sample and LiDAR data, using ANCOVA to justify this approach; (2) selection of independent variables; (3) regression model development; and (4) model assessment and validation. The information derived from the predictive LiDAR-AGB model at tree level should the information be available to natural resource managers to provide detail of forested areas that could be derived from the biomass assessment for improving management decisions.

- Chapter 5 (submitted for *peer-review, Forest Journal*): The overall accuracies obtained for individual tree detection over the two forest sites (84% for PFR site and 88% for FRIM site) indicate that the developed individual tree crown delineation framework can generate a map of multi-sized individual tree crowns in forests with accuracies comparable to visual interpretation. The dominant crown sizes of a forest can be automatically determined. The combined semi-variogram statistics and morphological analysis are valuable for the identification of dominant crown sizes of a forest. The semi-variogram statistics provides general information on tree crowns, i.e., the dominant features in a CHM image at the given spatial resolution, while the morphological analysis reveals local and detailed information on crown sizes. The density based algorithm approach is helpful to refine the problematic segments with more than one tree inside by determining the number of trees in the pre-identified segments based on the 3D LiDAR points and the rules set prior to knowledge of study areas.

- Chapter 6: Ordinary Least Square (OLS) regression as the parametric approach outperformed the non-parametric approach, Most Similar Neighbour (MSN) – based imputation, Gradient Nearest Neighbour (GNN) - based imputation and Random Forest (RF) as indicated by the performance measures; bias, RMSE and Adj-R^2 for predicting BA, V, Ht and AGB. The RMSE for OLS method after cross-validation for BA, V, Ht and AGB prediction were 1.40 m^2, 1.03m^3, 2.22 m and 96 Kg/tree. RF produced best overall results among the non-parametric methods tested. This study has shown the validation of the ITC delineation result from Chapter 5 and the capability of LiDAR data for predicting forest attributes at ITC level by comparing several statistical methods, both parametric and non-parametric in tropical forests. Studies in this forest type are few in number and less frequent than studies reported in other forest types such as temperate and boreal forests. As the best method varied according to response variables and performance measures, selecting a modelling technique should be carefully determined based on the objectives, conditions and scales at which the forest managers and researchers are interested.

7.2 Implications

These results have important implications for forest management and for the research community of forest departments, remote sensing, ecologist and geospatial analysists.

In support of sustainable forest management, up-to-date forest inventories are required to assess the composition, structure, and distribution of forest vegetation that, in turn, can be used as base information for management decisions that span across a range of spatial and temporal scales. Forest inventories typically report a representative estimate of height that is often defined as the average height of dominant and co-dominant trees (Gillis & Leckie 1993). To date, research and development activities have focused upon using LiDAR as a tool for characterizing vertical forest structure—primarily the estimation of tree and stand heights, with volume and biomass also of interest (Lim et al. 2003). With increasing availability of LiDAR data, forest managers have seen opportunities for using LiDAR to meet a wider range of forest inventory information needs (Nelson et al. 2003). For instance, height estimates generated from airborne remotely sensed LiDAR data were found to be of similar, or better accuracy than corresponding field-based estimates (Næsset & Økland 2002) and studies have demonstrated that the LiDAR measurement error for individual tree height (of a given species) is less than 1.0 m (Persson et al. 2002), and for plot-based estimates of maximum and mean canopy height with full canopy closure less than 0.5 m (Næsset 1997; Magnussen & Boudewyn 1998; Magnussen et al. 1999; Næsset 2002; Næsset & Økland 2002). Additional attributes, such as volume (Woods et al. 2008, 2011; Silva et al. 2016), biomass (Hyde et al. 2007; Gagliasso et al. 2014; Mauya et al. 2015), and crown closure (Bai et al. 2005), are also well characterized with LiDAR data (Means et al. 2000; Lim et al. 2003; Thomas et al. 2006; Wulder et al. 2012; White et al. 2013, 2016).

This thesis has concentrated on the use of LiDAR data, since the major goal of this thesis is to utilize LiDAR's unique potential for very high accuracy of the extraction of biophysical properties. LiDAR data alone may become an important forest inventory data source when combined with appropriate designed sample plots in the field and with appropriate modelling tools. The incremental integration of LiDAR-generated attributes into existing forest inventory data will, over the short term, promote the increased use of LiDAR for a range of forest applications, ultimately enabling time and cost savings for future implementations. The judicious and appropriate use of LiDAR data can enhance sustainable forest management practices by building upon existing knowledge and expertise in the forest management community.

Airborne LiDAR has emerged as one of the most promising remote sensing technologies, providing detailed, spatially explicit, three-dimensional information on forest structure, for operational applications in a wide range of disciplines related to the management of forest ecosystem (Vauhkonen et al. 2014). With increasing availability of LiDAR data, forest managers have seen opportunities for using LiDAR to meet a wider range of forest inventory information needs (Wulder et al. 2008). For instance, this thesis has shown promising results that LiDAR can be effectively used to extract individual tree crown structure and estimate AGB and other forest attributes with relatively high accuracy in the study site. With the successful of extracting individual tree crown structures and extract individual tree attributes including total height, crown height and crown diameter, this has led to a successful derivation of other bio-physical parameters e.g. V and AGB and being an important element to

describe the function and productivity of forest ecosystem (Dubayah & Drake 2000). This information is arguably of greatest interest to foresters (Lim et al. 2003). The accurate delineation of single trees within forested areas is of high value for a number of management tasks (Koch et al. 2006) and for obtaining better estimates of single tree utilizing LiDAR data, it is desirable to make consistent effort on algorithm development and testing. The LiDAR-AGB model can help forest managers to identify specific areas for management interventions, maintain biomass stock, optimize carbon sequestration and enhance forest productivity.

This study suggests that depending on what forest managers want to know about their forest, various modelling methods can be used to predict forest attributes of interest. Parametric method such as OLS, could offer a cost-effective way to determine AGB at a local scale. However, if multiple forest inventory variables are required, a non-parametric approach such as *k*NN could offer a better solution at more wider scales e.g. project level. This study also suggests that the use of LiDAR data as an explanatory variable in a regression model or in nearest neighbour imputation method can increase the accuracy of estimated AGB per hectare. The AGB model developed in this study could be transferred to be used to any similar forest types across the South-East (SE) Asia regions since the variety of vegetation species that exist in the two-forest site, PFR and FRIM is enough to represent any tropical forest available within this region (Fletcher et al. 2012). Furthermore, based on findings in chapter 4, the analysis of covariance (ANCOVA) test results for pooling data from different forest sites has proofed that forest environment does not give an impact on the LiDAR metrics, which means that the AGB estimation models developed in this study may be directly

transferred to other study areas within the similar tropical forest environment in SE Asia. Previous studies have shown promising result transferring AGB model between forest types by using power models based on mean canopy height (MCH) for example, Lefsky et al. (2002), found that 84% of AGB variation across three sites in North America can be explained using a model based on the squared MCH. Asner et al. (2012) found that, plot-scale biomass across four tropical forest sites in Peru, Panama, Madagascar, and Hawaii can be estimated by MCH LiDAR-based metrics after accounting for the differences of wood density and BA to MCH relationships for each site. More research is needed to assess model uncertainties.

Airborne LiDAR technology is developing at a rapid rate along with new applications. As technology standards continue to evolve and improve, data acquisition, processing approaches, parameters, and deliverables should remain fluid and be reassessed on a regular basis (Evans et al. 2009). Developing technologies such as; small footprint waveform LiDAR, single photon and flash sensors may provide alternatives to discrete return LiDAR. However discrete return LiDAR will continue to play an important role in natural resource decision making for the foreseeable future.

7.2.1 Special Implications on Forest Research Institute Malaysia (FRIM) and Forestry Department of Malaysia (FDPM)

FRIM and Forestry Department of Malaysia (FDPM) have a collaborative effort in conducting research and development in the field of forest and environment for the

betterment of forestry sector in Malaysia. FRIM and FDPM currently committed to four international climate change related projects from international donor agencies namely the Forestry and Forest Product Research Institute (FFPRI) Japan, the Asia Pacific Network for Sustainable Forest Management and Rehabilitation (APFNet) China, the Japan Aerospace Exploration Agency (JAXA), Japan and the International Tropical Timber Organisation (ITTO) (Frim.gov.my. 2018). Both agencies pledge their commitment to jointly implement the four projects which focus on Reducing Emissions from Deforestation and Forest Degradation (REDD+), biomass estimation, carbon stock assessment and the climate change issue. Out of the four projects, two of the projects are crossly related to the development of forest carbon monitoring methodologies and AGB and carbon stock quantification in the forest of Peninsular Malaysia by using Radar (L-Band ALOS Palsar and JERS-1) and LiDAR technologies (Frim.gov.my. 2018). This thesis will be a valuable guideline to FRIM and FDPM as it provides a crucial step-by-step on utilizing LiDAR for AGB modelling. Parametric and non-parametric approaches for modelling forest attributes as conducted in Chapter 6 seen as the most potentially beneficial for FRIM as the method could offer an option to conduct studies at local and project scale level depending on what forest managers want to know about their forest. Plus, to the best of author knowledge, non-parametric technique has never been implemented at local forestry project in Malaysia. Parametric method such as regression could be use in a cost-effective way if single piece of information is required, but if forest managers would like to know multiple forest inventory variables such as AGB, BA, and V, non-parametric method can be a good solution.

7.3 Future Considerations

Even though the proposed methods were demonstrated to be promising in terms improving the ITC delineation and modelling forest attributes at individual tree level, they can be further developed based on the following future considerations:

1. The number of individual trees collected at Pasoh forest reserve site and at FRIM site was limited, even though the author pooled the sample data from both forest site (N = 209), because of the labour-intensive ground measurement and time restriction. Additional tests with more validation datasets are needed to investigate the robustness and reliability of the methods, and comparison between the individual tree analysis and area based-approach would also recommend.
2. Adding supplementary data such as topographic variables (e.g. elevation and aspect which can drive temperature and moisture microclimates) to increase the chances of developing model at species level. Topographic variables are the good predictors of where some species occur. LiDAR data alone, using the only LiDAR derived predictor variables can be a source of potential to map stand structural attributes at the species level (Hudak et al. 2008, Dong 2009). Most of the species level mapping have used hyperspectral imagery (Hudak et al. 2008), often with upwards of 200 spectral bands, while LiDAR systems operate at a single wavelength. Within the spatial extent of a single two-dimensional image pixel, LiDAR surveys can provide literally hundreds of three-dimensional points. The canopy height profile derived from these LiDAR points could be termed a “structural signature” with at least comparable information content to a hyperspectral signature derived from an image pixel.

The study by Hudak et al. 2008 has demonstrated the plot level structural attributes of individual tree species can be simultaneously imputed from predictor variables derived solely from LiDAR data and conclude the high possibility of LiDAR potential to characterize detailed canopy structure variation and fine-scale topographic variation that constrains the distribution of species assemblages. With appropriate designed field sample plots in the field and with appropriate modelling tools, LiDAR alone may become an important forest inventory data source.

3. Individual tree analysis is still on the way to being more effective, efficient and accurate and this will certainly be benefited from the future advances in remote sensing technology such as full-waveform LiDAR (Gupta et al. 2010) and multispectral LiDAR (Lindberg & Holmgren 2017). Full waveform systems are very useful for capturing continuous vertical vegetation profiles and reconstructing tree height profiles since the full waveform systems record continuous reflected energy and can produce dense point clouds. With this great advance, it is expected to improve individual tree detection especially for one working in the dense forest structure like tropical trees (Zhen et al. 2016). The integration of multispectral data and discrete-return LiDAR is commonly used to improve tree species classification at tree level studies. Combination of multispectral information with 3D discrete-return data can lead to improvement in the accuracy of tree extraction and tree species classification as the advantages of both datasets can be fully utilized (Yu et al. 2017).
4. Accuracy of ITC is not only reliant on data quality and characteristics, but depends greatly on algorithm performance. Sophisticated advanced computer

vision algorithms such as mean shift clustering have been applied in individual tree delineation using LiDAR data (Ferraz et al. 2016; Hu et al. 2017). Mean shift clusters can indicate smaller or suppressed trees in the understory which is not possible to do using other kinds of methods (Amiri et al. 2016). This technique could be further explored as it believes can provide a promising way of producing relatively high accuracy of detecting smaller trees in the understory better than CHM-based methods (Hu et al. 2017).

APPENDIX 1 – Journal paper of Chapter 4

PAPER I

Journal of Tropical Forest Science 29(4): 465–484 (2017)
https://doi.org/10.26525/jtfs2017.29.4.465484

MODELLING INDIVIDUAL TREE ABOVEGROUND BIOMASS USING DISCRETE RETURN LIDAR IN LOWLAND DIPTEROCARP FOREST OF MALAYSIA

WS Wan-Mohd-Jaafar[1, *], IH Woodhouse[1], CA Silva[2], H Omar[3] & AT Hudak[4]

[1]*School of Geosciences, University of Edinburgh, Drummond Street, EH8 9XL, Edinburgh, United Kingdom*
[2]*Department of Natural Resources and Society, University of Idaho, 875 Perimeter Drive, Moscow, ID 83844, United States*
[3]*Forest Research Institute Malaysia, 52109 FRIM, Kepong, Selangor, Malaysia*
[4]*US Forest Service (USDA), Rocky Mountain Research Station, RMRS, 1221 South Main Street, Moscow, Idaho, USA – 83843*

**W.B.Wan-Mohd-Jaafar@sms.ed.ac.uk*

Submitted November 2016; accepted March 2017

Light Detection and Ranging (LiDAR) has become a common means for predicting key forest structural attributes. The aim of this study was to explore the relationship between individual tree LiDAR-based metrics and field data on tree attributes from a tropical rainforest in Peninsular Malaysia, to assess the correlation between LiDAR and field data at the individual-tree level for aboveground biomass (AGB) estimates. The model was developed using multiple regression analysis, with a non-linear power model being used to fit the predictive models. The AGB model was developed based on estimated AGB at the field site and LiDAR data using the following methodology; (1) pooling of both field sample and LiDAR data, using ANCOVA to justify this approach, (2) selection of independent variables, (3) regression model development and (4) model assessment and validation. LiDAR height percentile (h80) and crown width (CW) measurement were found to best fit the data as evidenced by Adj-R^2 value of 0.63, root mean square error (RMSE) of the model of 14.8% and analysis of the residuals. This study provides an analytic framework for developing a predictive LiDAR-AGB model at tree level as LiDAR derived information helps natural resource managers to provide details of forest that could be derived from the biomass assessment to improve management decisions.

Keywords: LiDAR-AGB model, tree level, tropical rainforest, height percentile, crown width

INTRODUCTION

Tropical forests in Southeast Asia have declined acutely over the past several decades (William 2007). In particular, according to a new global forest map in partnership with Google, Malaysia had the world's highest rate of forest loss between 2000 and 2012 (Butler 2013). Reducing emissions from deforestation and forest degradation (REDD+) is the framework for conserving and enhancing carbon stocks of forested area in the tropics (UNFCC 2007). For REDD+ implementation, accurate estimation and monitoring of carbon stocks are required at the national and subnational levels. To establish robust and transparent monitoring systems, a combination of ground-based sampling and remote sensing approaches was recommended (UNFCCC 2009). Aboveground biomass (AGB) of trees in tropical forests account for significant part of the total carbon pool (Houghton et al. 2001). Therefore, estimating AGB is critical to accurately quantifying carbon stocks in the tropics (Gibbs et al. 2007).

Tropical forests are known for their complex stand structure and abundant diversity in species composition (Steininger 2000) and estimating AGB from remote sensing data in this dense forest is challenging. Satellite-mounted optical sensors have been widely used to estimate AGB (Anaya et al. 2009). However, optical sensors acquire information from the upper canopy and are unable to measure the three-dimensional structure, including canopy height and sub-canopy topography (Lu 2006), which limits their utility to quantify AGB in tropical forests with complex canopy structures. Radar sensors (e.g., ALOS/PALSAR) use active microwave signals to generate an image, and these can be used to determine forest vertical structure (Gibbs et al. 2007), even in areas of high cloud cover such as the tropics. Radar sensors can be used for

relatively young or homogeneous forests, but their accuracy and sensitivity decrease in old-growth forests unless longer wavelengths are used (Hamdan et al. 2015).

Light Detection and Ranging (LiDAR) emits laser pulses and measures the return time of scattered returns to directly estimate the height and vertical structure of forests (Dubayah & Drake 2000, Lefsky et al. 2002). LiDAR can be acquired at high sampling density with excellent geometric accuracy and reveal AGB variation at fine spatial scales (Reutebuch et al. 2005, Mallet & Bretar 2009). LiDAR is therefore well placed to bridge the scale gap between satellite observations and field measurements (Asner 2009).

LiDAR remote sensing systems can be distinguished based on the way in which return signals are recorded (discrete return or waveform), scanning pattern (profiling or scanning), platforms (airborne, spaceborne or ground based) and footprint sizes. The most common configurations of LiDAR systems is airborne small footprint discrete return scanning LiDAR, as used in this study. Airborne discrete return LiDAR has been used in a large number of studies for mapping biomass mainly using two approaches: (i) area-based and (ii) individual tree-based methods.

In this study, the individual tree-based method was considered at Pasoh Forest Reserve in Peninsular Malaysia. This site contains mixed species and is dominated by trees from the Dipterocarpaceae family, which is common in lowland dipterocarp forest. Lowland dipterocarp forest is one of the most species-rich communities in the world, with more than 200 species per hectare (Symington 1943, Wyatt-Smith 1964). Individual tree detection is seen as the most relevant approach to extract tree structural attributes in tropical rainforest characterised by a complex three-dimensional structure. LiDAR forest inventory methodologies based on individual tree detection have been widely studied, but are not widely used in practice, due to the difficulties of tree detection in various forest conditions, especially in dense, closed-canopy tropical forests (Kaartinen et al. 2012).

The most widely used LiDAR metrics for AGB prediction are various height metrics that are associated with field measurements through empirical models (Kaartinen et al. 2012). LiDAR metrics can be calculated based on first return, last return or all of the returns (Qi 2013). In this study all returns were used to maximise the information content. Unlike most of the published algorithms that detect individual trees from a LiDAR-derived raster Canopy Height Model (CHM), this study worked directly with the LiDAR point cloud data and field data to distinguish individual trees and to estimate individual tree metrics. The CHM is a raster image interpolated from LiDAR points depicting the top of the vegetation canopy (Khosravipour et al. 2014). As a result, the CHM can have inherent errors and uncertainties from a number of sources (Khosravipour et al. 2014). However, by directly interpolating the raw LiDAR point cloud to extract individual trees, the measurements are not affected by the errors associated with interpolation, and the important 3D forest parameters can be extracted directly from the LiDAR returns that make up each tree (Li et al. 2012). To our knowledge, the use of direct LiDAR point cloud to detect individual trees and extraction of LiDAR height metrics in tropical rainforest of South East Asia has been little studied and this is one of the first studies to implement this approach for individual tree LiDAR-AGB modelling in Malaysia.

Due to the structural complexity of tropical rainforests, further research is needed to identify the relationship between AGB measured in the field and LiDAR height metrics, and to determine how these relationships impact the accuracy of predictive models. This research integrates, tree-level field-sample data with LiDAR variables to predict AGB in tropical rainforest. The goal of this study was to model individual tree AGB based on trees that were mapped in the field, with the intention that the model could later be applied to a wider area. The immediate objectives were to: (1) develop a non-linear AGB model based on field sample plot and LiDAR data and (2) validate the model in terms of accuracy and precision.

MATERIALS AND METHODS

Site description

The Pasoh Forest Reserve (PFR) in Peninsular Malaysia was selected as the study area for this research because of its species and structural diversity. The PFR study site (2.98 N 102.31 E) is located about 8 km from Simpang Pertang, Negeri

Sembilan. PFR has an area of approximately 140 km^2, and is mainly covered with lowland dipterocarp forest, with upland dipterocarp forests near the north-eastern boundary. The core area of old growth forest is approximately 600 ha. Most of the surrounding area has been logged in the past, resulting in several areas of regenerating lowland forest. The PFR is one of the most species-rich forest communities in the world with 340,000 trees, ≥ 1 cm diameter at breast height (DBH), consisting of 818 species. The field measurements for this study were collected within the 6-ha International Biological Programme (IBP) plot (known as the 'ecological plot') in which all trees, ≥ 5cm in DBH, have been measured and mapped since the early 1970s.

In order to represent a wide sampling of lowland dipterocarp forest, the LiDAR-AGB model also incorporated field-sampled tree data recorded from the Forest Research Institute Malaysia (FRIM) Forest Reserve. FRIM is located at Kepong, Selangor, approximately 16 km north-west of the capital city of Kuala Lumpur and 140 km away from the main test site of PFR. This 600 ha tropical forest contains approximately 15,000 species of plants. This forest has a similar structure to the study plot at PFR, where the area was stripped of its original forest cover and logged over 100 years prior, providing a generation almost as old as the primary forest. Both forest study sites are categorised as lowland forest (Figure 1).

Field data collection

The data used in this study include a vegetation field sample data collected in 2014 and LiDAR data collected in 2012 at PFR, as well as the

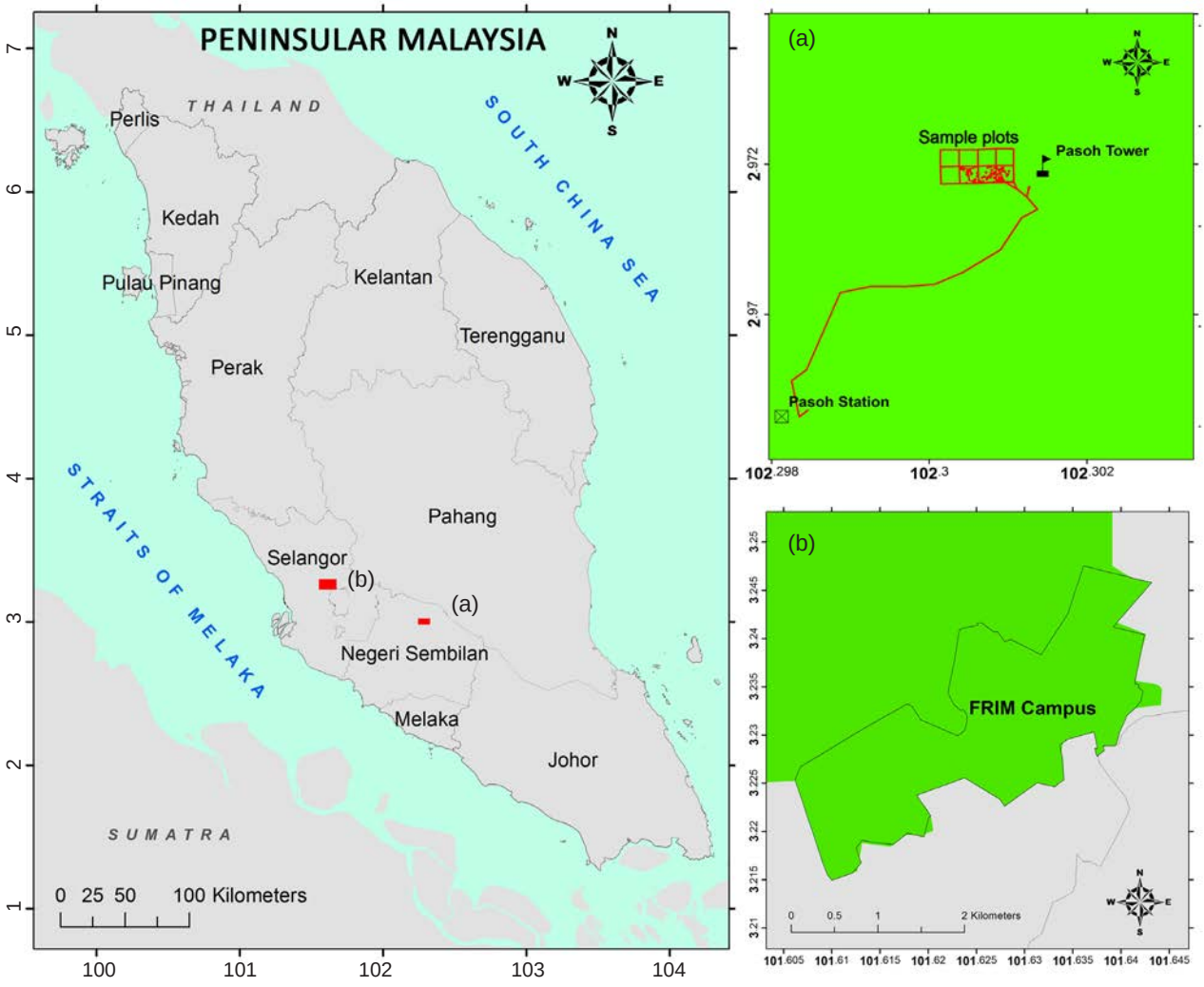


Figure 1 Study area (a) PFR, 6-ha IBP plot location and (b) 600-ha FRIM located inside FRIM campus

vegetation field sample data and LiDAR data in FRIM both collected in 2014.

Fieldwork was conducted at PFR, Negeri Sembilan, Malaysia from 28th to 30th October 2014. Six main forest parameters were collected; horizontal position (x, y) of individual trees, DBH, total height, bole height, crown diameter and tree species. Stratified random sampling of a rectangular area of 50 m × 100 m (0.5 ha) was split into 8 experimental plots with dimension of 25 m × 25 m (Figure 2.). Each trees with DBH > 10 cm was measured within plots A1, A2, A3 and A4. The criteria was changed for plot B1 to B4 such that in these plots only emergent trees with height > 20 m were measured. The measurement strategy was designed in such a way as to facilitate the mobility of sampling work in the field. A total of 105 individual trees were used for final assessment.

Considering that the distribution of the AGB sample data deviates far from a normal distribution, the sample size of the selected 105 trees is possibly too small to build a reliable regression model representative of all species across both study areas. To address this issue, the field sample acquired in early 2014 from FRIM were evaluated to determine whether it could be incorporated into the regression model. Since both PFR and FRIM are under the authority of FRIM, a similar sampling method and plant measurement protocol was applied at FRIM as at PFR. The FRIM is similar to PFR in terms of topography, vegetation composition and structure and land use history, because both forest are managed under the authority of FRIM. Some environmental and successional differences exist between PFR and FRIM. However, the AGB characteristics are likely to have more important impact on the LiDAR metrics than the environmental factors, assuming a relationship exists between the AGB and the LiDAR variables. To test this assumption, an analysis of covariance (ANCOVA) was conducted in this study as detailed in section ANCOVA.

The trees locations were determined using geographic coordinates of the plot centres, and the direction and distance of trees, relative to the plot centre. The plot centres were measured with a handheld global positioning system (GPS) device and the locations were post processed with local base station data, resulting in an average error of approximately 0.5 m horizontally. Tree heights were measured using a hypsometer and the DBHs and crown diameter were measured using a diameter tape (d-tape). Crown diameter was measured in four cardinal directions with

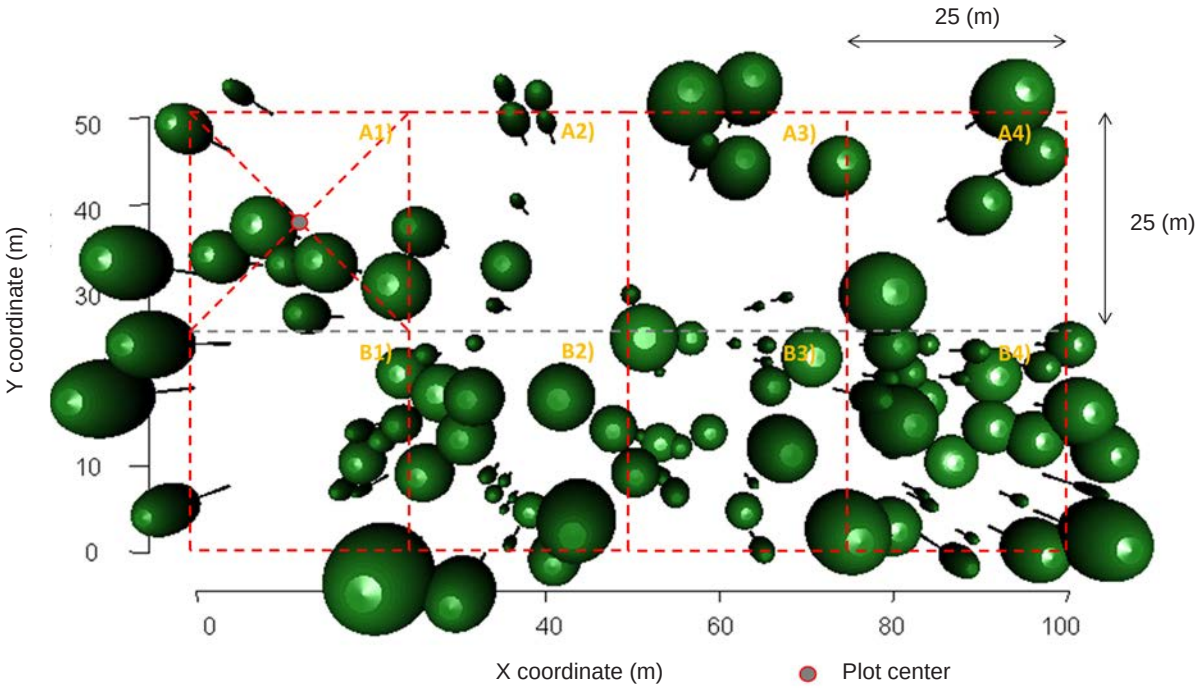


Figure 2 Dimensions of plots created within 0.5-ha plot in PFR

respect to tree trunk. Statistics describing the trees are in Table 1.

Aboveground biomass (AGB) estimation

An abundance of allometric models to calculate AGB have been developed in South East Asian tropical, secondary and Dipterocarp forests (Basuki et al. 2009, Niiyama et al. 2010). The selection of allometric equation for AGB are dependent on the characteristics and composition of the study area. In this study allometric equations for calculating AGB from field measurement were selected from Chave et al. 2014. This is a pan-tropical multispecies allometric equation whereby the study site is one of the many test sites of the Center for Tropical Forest Science (CTFS), used to develop the allometry model. Moreover, the chosen equation is a tree parameter incorporating DBH, height and wood density as predictor variables, describing the shape of the tree. Wood density is an important determinant of AGB, especially when a broad range of vegetation type is considered, and this model has been improved by including twice the number of trees as previous studies (Chave et al. 2014). A careful selection of allometric equations is important to reduce the uncertainty in estimating AGB:

$$AGB = 0.0673 \times (\rho D^2 H)^{0.976} \quad (1)$$

where ρ = wood density (g/cm^3), D = diameter at breast height (DBH) (cm) and H = tree height (m).

LiDAR operations

The LiDAR data were obtained from a private Malaysian airborne company, using an IGI LiteMapper-5600 system with a Riegl Q560 LiDAR sensor scanning at a ± 22.5^0, at a line rate of 60 line s^{-1}. Data processing steps include the production of radiometrically calibrated data (level 1), traceable to national standards for derived geophysical data products (level 2), which followed the application of an atmospheric correction. Finally, data were geometrically rectified to the local geo-reference co-ordinate system with user-defined ground control points (level 3). The data supplied were checked for quality, and delivered as classified and unclassified point clouds in both ASCII XYZ and LAS formats with a projection of UTM_Zone_48N. The data have been validated and quality checked and any possible low points have been removed. The root mean square (RMS) achieved by specific land class of forest was 0.092 m, and proved that the accuracy of LiDAR data is within tolerance of 0.15 m in vertical offset. The average point density was 8.8 points m^{-2}. The LiDAR data for FRIM was collected by FRIM. The data supplied was the same standard as the data received from PFR, with the average point density 7.5 points per square meter. Specification of the LiDAR system used at PFR and FRIM are summarized in Table 2.

Data post-processing

FUSION software was used to process the LiDAR data to generate three main products: the digital

Table 1 Characteristics of individual tree field samples from both sites

Forest site	Parameters	Min	Max	Mean	SD
PFR	Tree Height (m)	7.30	39.90	18.16	6.87
FRIM		12.90	49.60	38.93	9.36
PFR	DBH (cm)	10.10	82.30	29.84	19.33
FRIM		10.00	82.20	37.92	20.86
PFR	Volume (m^3)	0.02	8.00	1.03	1.70
FRIM		0.12	9.14	2.46	2.47
PFR	AGB (kg)	8.70	8067.16	918.74	2219.91
FRIM		122.1	7771.90	2421.80	2162.92

PFR = Pasoh Forest Reserrve, FRIM = Forest Research Institute Malaysia, DBH = diameter at breast height, AGB = aboveground biomass, SD = standard deviation

Table 2 LiDAR acquisition parameters

LiDAR sensor	IGI Lite Mapper-5600 Riegl Q560	ALTM Gemini laser system
Pulse Rate	Range between 70 KHz to 240 KHz	70 KHz
Scan Angle	± 22.5^0	± 25^0 , in increments of ± 1^0
Scan Pattern	Regular	Regular
Effective rate	46,667 Hz	33- 167 KHz
Line s^{-1}	Max 160	Max 160
Flying height	700-1000m	150- 4000 m nominal
Laser points m^{-2}	8.8 points m^{-2}	7.5 points m^{-2}
Max Above Ground Level	1040m (3411ft)	5000 m (16404ft)
Data Format	ASCII XYZ and LiDAR exchange format (LAS)	LiDAR exchange format (LAS)

terrain model (DTM), the digital surface model (DSM) and the canopy height model (CHM). The 'catalog' function was used to evaluate the LiDAR characteristics.

A digital terrain model (DTM) was created in two steps from the discrete return LiDAR data: (1) the data were filtered to remove the above-ground returns using algorithm an adapted from Kraus and Pfeifer (2001), performed by using the 'groundfilter' function and (2) the DTM was created by calculating the average elevation from the remaining (ground) LiDAR returns within a cell (cells that contain no points were filled by interpolation using neighbouring cells) by using function 'gridsurfacecreate'. The digital surface model (DSM) which represents the earth's surface and includes the trees and other objects on it, was created using 'canopymodel'. The discrete return point clouds were then normalised against the ground surface height and extracted for each plot using the coordinates of the lower left and upper right plot corners by using the 'clipdata' function. After heights were normalised, the canopy height model (CHM), which represents the height of the forest was generated by using 'canopymodel' function. The CHM was created for visualisation and image interpretation purposes and for manual co-registration between field sample data and LiDAR data, but not for LiDAR metrics extraction. Rather, individual tree LiDAR metrics were extracted directly from the normalised point cloud data. All point cloud data processing was performed using FUSION software (McGaughey 2014).

Individual tree extraction—co-registering LiDAR and field sample data

The development of the co-registration procedure was based on field sample data collected during fieldwork. The attributes relevant for this study are given in Table 3. The position of the plot edges were georeferenced with a total station. Coordinates of each tree were determined by manually seeking the optimum fit between tree positions and heights measured by forest sampling and CHM. To do this, the absolute positions of the trees within each plot were calculated from the geographical coordinates of the sample plot centres and the coordinates of the individual trees measured from field. To obtain an interpretable best fit of the tree pattern, these coordinates were then converted into ArcGIS shapefiles, which, in combination with the field height of each tree and crown diameter measured in the field, formed a polygon representing the crown dimensions of each tree. Each of the tree crown polygons was assigned a unique identification number (ID) and projected to the same projection as the LiDAR (UTM_Zone_48N).

To facilitate visual comparison with the LiDAR CHM, the field tree crown polygon was visualised and manually moved to best fit the shape and height of the CHM. Errors of this manual co-registration method are expected to lie at the subpixel level (i.e. < 1.0 m). Out of the 142 individual trees, 3 trees were initially removed due to the condition of being broken at first branch, which left 139 individual trees to be assessed manually. From these, 105 could be unambiguously manually co-registered. After manual co-registration, the estimated absolute planimetric accuracy of the tree location was ± 0.5 m. After carefully co-registering the individual tree polygons with the LiDAR data, using the X, Y coordinates and the crown diameter measurement for each individual tree, the tree crown polygons were used to clip the LiDAR point cloud data such that the points within each

polygon clouds were assigned the same ID as the individual tree crown polygon ID. Individual tree LiDAR metrics were then computed using the rMetrics function in the rLiDAR package (Silva et al. 2015). The generated metrics from LiDAR were used to model the individual tree heights represented by the LiDAR height metrics (Figure 3).

LiDAR metrics

Discrete return LiDAR metrics are descriptive structure statistics, calculated from the measurement of height normalised point cloud in three-dimensional space (Lefsky et al. 2005). In this study, 30 metrics for each individual tree were calculated including: (1) selected percentile heights (i.e., 5,10,….99 denoted as h_5,h_{10},….,h_{99}, respectively), maximum height (h_{max}), crown base height (CBH), mean height (h_{mean}), median height (h_{med}), mode height (h_{mode}) and (2) variability of height measures i.e., coefficient of variations of height (h_{cv}), variance (h_{var}), kurtosis ($h_{kurtosis}$), skewness ($h_{skewness}$) and standard deviation (h_{sd}). The CW was computed using the chullLiDAR 2D function after computing the canopy area. A summary of the metrics with corresponding descriptions is shown in Table 3. To examine the impact of laser returns on AGB estimation, LiDAR metrics point clouds based on all returns were generated.

Statistical analysis

Assumption of homogeneity of regression slopes in analysis of covariance (ANCOVA) was tested before developing the full model. This was to ensure that field sample data from FRIM

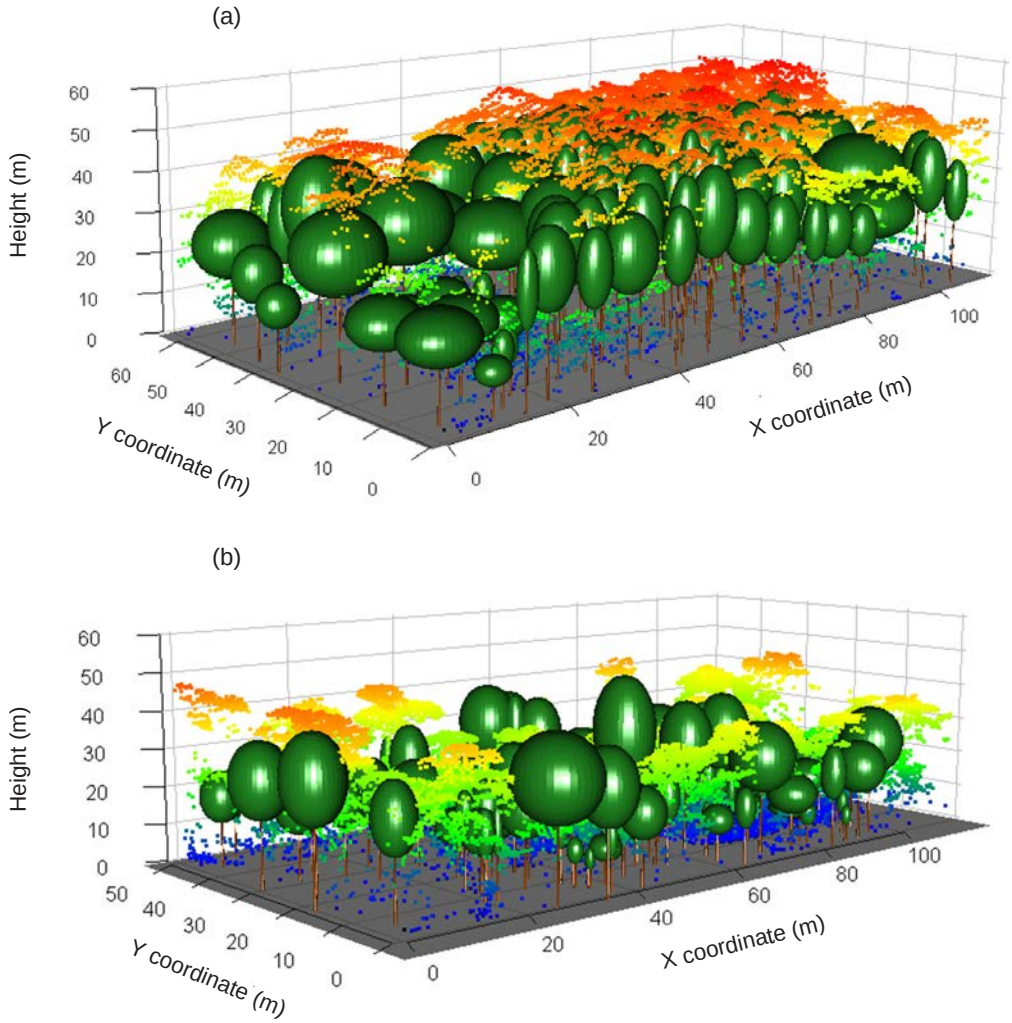


Figure 3 3-dimensional tree height model; individual tree crowns extracted from the LiDAR point clous (overlaid) at (a) FRIM and (b) PFR; trees measured in Pasoh site were based on selected criteria, not all trees were measured, thus, some trees were represented by point cloud

Table 3 List of LiDAR metrics

Metrics	Description
Total number of returns	Total number of discrete LiDAR measurements for individual tree
Crown base height (CBH)	Points height above the crown base height
Maximum height (hmax)	Maximum height above ground of all LiDAR returns for individual tree
Mean height (hmean)	Mean height above ground of all LiDAR returns for individual tree
Median height (hmed)	Median height above ground of all LiDAR returns for individual tree
Mode height (hmode)	Mode height above ground of all LiDAR returns for individual tree
Standard deviation (hsd)	Standard deviation of heights of LiDAR returns for individual tree
Percentile height (h5, h10,h20,h25,h30,h40,h50, H60,h70,h75,h80,h90,h95,h99)	The percentiles of the canopy height distributions (5th,…,99th) of all returns
hcv	Coefficient of variation of heights of all LiDAR returns
hvar	Distribution of average height variance derived from all LiDAR returns
hkurtosis	Distribution of height kurtosis
hskewness	Distribution of height skewness
CW	Crown width

could be pooled with PFR for regression model development. To estimate the AGB in PFR, regression model that represents the relationship between the selected LiDAR metrics and the AGB was developed. The statistical analysis involved three steps: (1) selection of independent variables, (2) regression model development and (3) model assessment.

ANCOVA

Since AGB at FRIM differed from PFR due to the different stand conditions, it is important that data pooling does not compromise the LiDAR-AGB relationship for the regression model in the study area. This can be tested using an ANCOVA approach. The AGB was estimated using the selected appropriate allometric equations and field data. Initially a linear regression model was chosen to observe any non-random patterns in the residuals. A curve was fitted to the trends of residuals for both forests.

ANCOVA was conducted with a general regression model having one continuous outcome variable and one or more factor variables. It tests whether certain factors have an effect on the outcome variable after adjusting for the effects of confounding factors. In this study, ANCOVA was used to determine whether the sampled data from PFR and FRIM forest could be pooled for the AGB regression model in representing both study areas through testing the assumption of homogeneous regression slopes. This assumed that the relationship between the AGB and the LiDAR metrics in PFR and FRIM were independent of the regional conditions. To test this assumption, interaction effects were added into the regression model by adding the product of covariates and the regional factor. A power transformation was used to test this assumption because it best fit the sample data in the study:

$$Y = (\beta_0 + \beta_{site}) + (\beta_1 + \beta_{1site}) \text{ Ln CW} + (\beta_2 + \beta_{2site}) \text{ Ln (h80)} \quad (2)$$

where Y = field values of AGB (kg tree^{-1}); site is a dummy variable for representing the different site (1 = FRIM, 0 = PASOH) and CW and h80 is the independent variables. When site is zero, the intercept and the slopes become β_0, β_1 and β_2 respectively in the model, which is derived only from PFR. When site is 1, the intercept and the slopes of the model include the addition of field sample data from FRIM. This approach was used to determine whether adding the sampled data from FRIM would impact the intercept and the slopes of the model significantly. The null hypothesis for the F test that $\beta_0 = 0$, $\beta_1 = 0$ and $\beta_2 = 0$; adding the FRIM data changed neither the intercept nor the slopes.

Selection of independent variables

The strength of relationships between AGB and the 30 LiDAR metrics were tested with the coefficient of determination (R^2). Regression

models with highly correlated independent variables are not stable from a statistical perspective and are hard to interpret from a biological perspective (Li et al. 2008). Following the variables selection method of Næsset et al. (2005) who minimised the number of LiDAR metrics to avoid information redundancy and promote parsimony, the original LiDAR metrics were reduced to non-correlated principle components. It is difficult to interpret the principle components themselves because they are in linear combinations of the original LiDAR metrics and do not themselves have a clear physical meaning (Li et al. 2008). Some of the metrics selected introduced serious multicollinearity with VIF value greater than 3. The solution was to shift to the variable selection method of (Næsset & Gobakken, 2008) where all of the independent variables were included as possible predictor variables for selection using both stepwise variable selection with an R^2 improvement technique.

The model simulates the relationship between field AGB and LiDAR variables in all sites and in combination. The estimated AGB, a dependent variable, was calculated based on the converted ground measurements taken during October, 2014 at PFR and April, 2014 at FRIM. All 30 variables of LiDAR metrics were used as potential model-independent variables.

Regression model development

Power functions have been used in a large number of studies for AGB estimation, probably because most allometric equations for calculating AGB in the field are power functions. The relationship between AGB and forest height has been well described (Morel et al. 2011, Wang et al. 2013), and the power function is widely accepted (Niklas & Enquist 2001) as:

$$B = a\,(H)^{b} \quad (3)$$

where B = plot AGB (mg ha^{-1}), H = field tree height (m) and *a* & *b* = coefficients.

The allometric equations used to derive AGB from field data in this study were based on a power function developed by Chave et al. (2014) and thus, provide a solid justification for the chosen model. Since most allometric equations for calculating tree-level AGB from field measurements are power models, AGB and LiDAR metrics were log transformed to fit a regression model. This is to reduce the heterogeneity of the regression residual variance. A multiplicative model formulated as equation (3) was selected, which can be translated into a linear form according to equation (4). This type of model has been used successfully by others to model various forest biophysical properties (Naesset 2002, Lim et al. 2003). The natural logarithm transformations required in equation (4) also ensure that, in most cases, regression assumptions are not violated. In the regression analysis, a linear multiplicative model was used which correspond to multiplicative power transformation, as used successfully by others to estimate various forest attributes. All derived variables were transformed to the natural logarithm. First, AGB was regressed against height metrics and AGB as a dependent variable. Multiple linear regression analysis using all independent variables was then carried out for both sites independently and in combination. Stepwise selection using Akaike Information Criterion (AIC) was performed in R to select variables included in the final model. The AIC is a goodness of fit measure that favours smaller residual error in the model, but penalises the inclusion of predictors and helps avoiding overfitting. At each step, individual variables were either added or deleted and the next model with lowest AIC was retained in the next step. The coefficient of determination (R^2), the adjusted coefficient of determination (Adj-R^2) and the root-mean-square error (RMSE) were calculated for model comparisons.

$$Y = \beta 0 + \beta_1 CW + \beta_2 h_{max} + \beta_3 CBH + \beta_4 h_5 + \ldots\ldots + \beta_{30} h_{99} \quad (4)$$

$$\ln Y = \beta_0 + \beta_1 \ln CW + \beta_2 \ln h_{max} + \beta_3 \ln CBH + \beta_4 \ln h_5 + \ldots.. + \beta_{30} \ln h_{99} \quad (5)$$

where Y = field values of AGB (mg ha^{-1}), h_{max} = maximum height of canopy, CBH = crown base height, h5,…, h99 = percentiles corresponding to 5,…, 99% of the laser canopy height (m), CW = crown size width. Both stepwise variable selection and the maximum R^2 variable selection techniques were applied to select LiDAR variables, to be included in the models. When using stepwise selection method, no independent variables were left in the models with a partial F statistic significance level greater than 0.05. The

best fitting models were selected based on the lowest AIC value. The variance inflation factor (VIF) was used to address multicollinearity issues by calculating and monitoring the size of the condition number. Because the best models suggested by the stepwise procedure might have many independent variables that could introduce multicollinearity problems, the maximum R^2 improvement technique searches for the best one-variable model.

In accordance with the objectives of this study, the influence of forest site on the estimated AGB models were assessed by extending the preliminary regression models derived above with dummy variables representing these factors. Since both sites were lowland dipterocarp forest with almost similar species, the homogeneity test to assess whether both datasets can be pooled for model development was only assessed on variables related to site properties. To assess the effects of different forest site, the dummy variables were assigned a value of 0 for sampled data in Pasoh and a value of 1 for sampled data in FRIM. The result are discussed in ANCOVA analysis in ANCOVA.

Model assessment

Two questions were essential in assessing the model: (a) how well does the model fit the sampled data (model fitting analysis) and (b) is the model generalisable outside the sampled data? In the model fitting analysis, potentially influential data outliers were identified using the studentised deleted residual and Cook's distance statistic to assess a data point's influence on the regression coefficients of the regression model; generally, it should be considered a potentially influential point when its value exceeds 1. The goodness of fit of the model was evaluated using an adj-R^2 and the RMSE, and standardised residuals were used to check the following model assumptions: (1) the equal variance for all independent data and (2) no systematic pattern between the regression model residuals and the predictions. The R^2 evaluates the fit of a linear regression model to the sampled data. However, the impact of the degree of freedom on the model accuracy is well known; the more predictive variables in the model, the higher the R^2. Clearly, the R^2 is not a good index to assess the model's goodness of fit. An Adj-R^2 removes the impact of the degree of freedom, and thus provides a more conservative measure of the model's goodness of fit.

The assumption of normality of error terms were assessed using the Shapiro-Wilk test (Shapiro & Wilk 1965) and heteroscedasticity were tested using the Breusch-Pagan test (Breusch & Pagan 1979). Comparisons between models were based on their predictive capabilities with respect to the coefficient of determination (R^2), root mean square error (RMSE) and AIC between observed and predicted values. Once the best model was chosen, 10-fold cross validation was performed. RMSE were used as a way to evaluate the measured values and to estimate the true value. RMSE was calculated as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (yi - \hat{y}i)^2}{n}} \qquad (6)$$

where y_i = observed values and $\hat{y}_i$ = predicted values for the *i*th compound, respectively, and *n* = number of samples in the training set.

RESULTS

A linear regression model was chosen to evaluate whether a relationship exists between AGB and LiDAR variables. Residuals versus modeled AGB were plotted to examine the assumption of linearity. To help observe patterns of residuals, a curve was fitted to the trends of residuals (Figure 4). The plot clearly displayed non-linear trends in the residuals.

Independent variables selection

Regression models with log-transformed variables were estimated with LiDAR derived metrics as the only independent variables. The models were selected by a stepwise regression procedure using all possible independent variables from both sites independently and combined. These are summarised in Table 4. To simplify the models, the maximum R^2 improvement technique was used to find the best variable combination and the best-two variable model and so-forth, with the benchmark of the best fitting model being the one with the lowest AIC value. Multicollinearity was further evaluated to confirm that all independent variables had correlations below 0.90, a VIF value below 5, and high correlation with AGB.

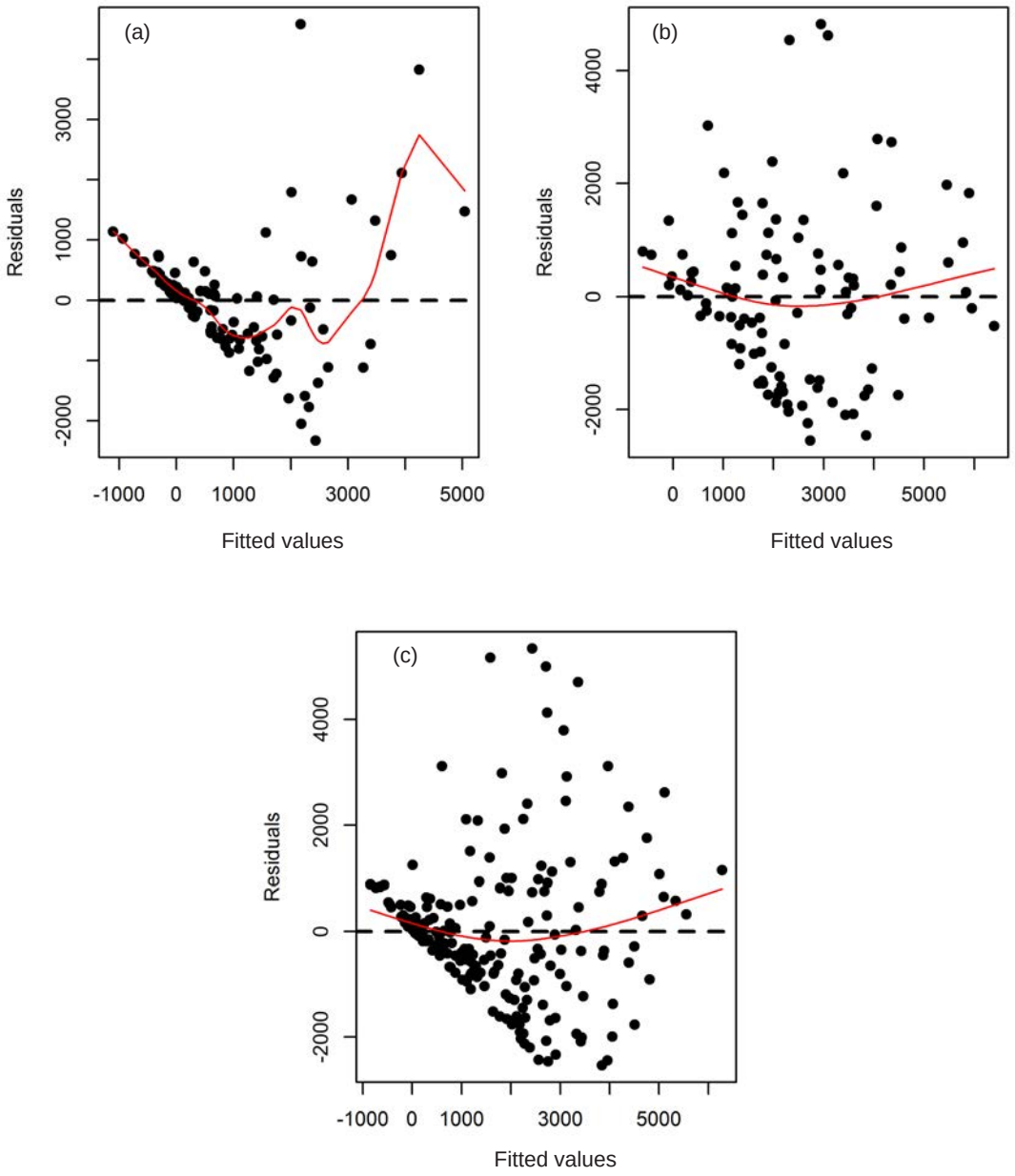


Figure 4 Residuals on fitted linear regression, (a) PFR site, (b) FRIM site and (c) pooled data between PFR and FFR site, all indicating randomness existed in their relationship

Based on high Adj-R^2 and low rRMSE values, the predictive models, combining data from two forest sites, proved as an efficient strategy to predict AGB of the region. The best model was selected based on lowest AIC value, however, the selection of the best variables was determined by the correlation coefficients and p-values of the partial F statistic of selected metrics. Figure 5 shows a scatterplot matrix of the two models with lowest AIC value. The model combining four variables (CW, CBH, h5 and h75) gave the lowest AIC value and highest Adj-R^2, but this combination introduced multicollinearity with VIF value > 5 and high inter-correlated independent variables with r > 0.90. This value can be summarised through Figure 5(a). The model combining the CW and h80 variables, however, passes all the diagnostic tests with VIF < 3 and low correlation between predictors with r < 0.70 used for further analysis in this study, as summarised in Figure 5(b). The histogram of each metric are drawn in the diagonal line. The kernel density overlaid and the significant asterisks shows the level of significant (*0.05,

Table 4 Top four best variable combinations, based on lowest AIC values from two forest sites considered independently and in combination

Study site and model		Selected independent variables (natural log transformed)	Adj-R^2	AIC	RMSE (kg tree^{-1})	rRMSE
PFR	1)	CW, CBH, h70	0.74	256.45	0.79	13.65%
	2)	CW, h70	0.74	256.72	0.79	13.74%
	3)	CW, CBH, h5, h70	0.74	257.05	0.78	13.61%
	4)	h90	0.61	298.71	0.97	16.84%
FRIM	1)	CBH, hmode	0.25	300.10	1.00	13.73%
	2)	CW, hvar, hcv, h90	0.22	305.85	1.01	13.96%
	3)	h90	0.11	316.81	1.09	14.96%
	4)	hmedian	0.12	317.70	1.08	14.86%
Pooled Data site	1)	CW, CBH, h5, h75	0.63	579.65	0.95	14.52%
	2)	CW, h80	0.63	581.90	0.96	14.68%
	3)	CW, CBH, h5	0.63	591.15	0.97	14.96%
	4)	h90	0.53	627.68	1.12	16.41%

PFR = Pasoh Forest Reserrve, FRIM = Forest Research Institute Malaysia, AIC = akaike information criterion CW = crown width, RMSE = root mean square error, CBH = crown base height, h = percentiles

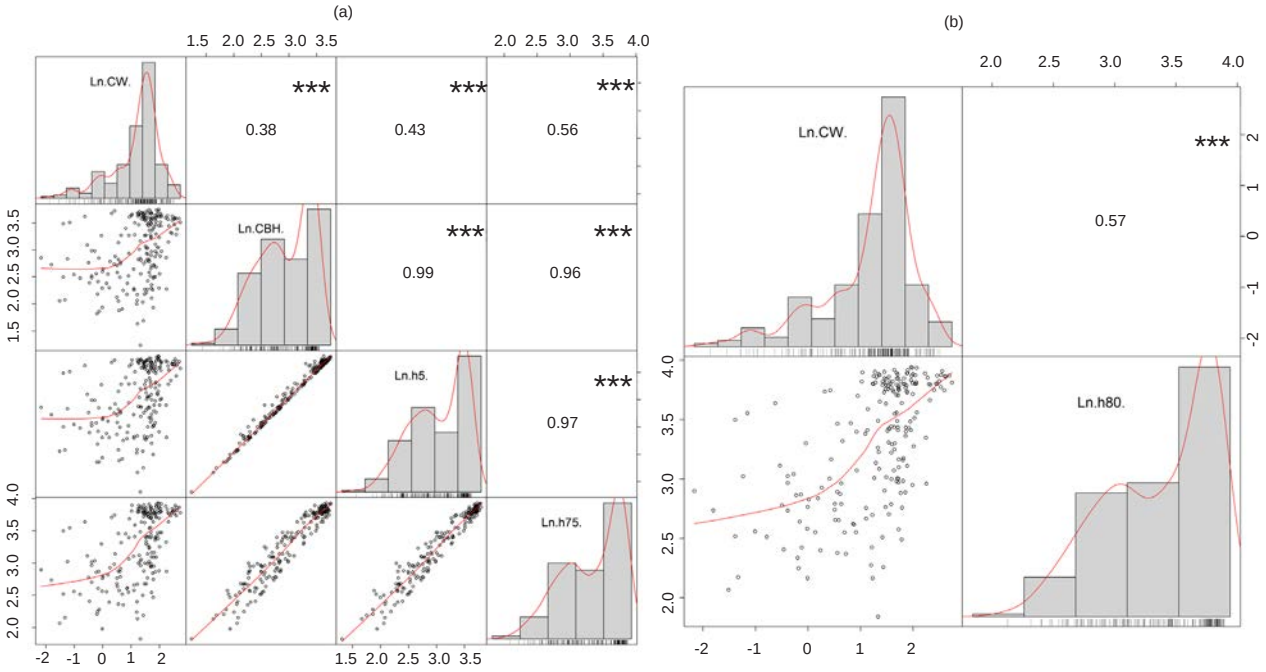


Figure 5 The matrix of scatter plots (lower panel) and the correlation coefficients (r) (upper panel) for all detected trees from all possible best metrics in pooled data site; the correlations are presented as (a) high correlation and (b) low correlation

**0.01 and ***0.001). Further model assessment will focus on the model based on the combined sites instead of the individual test sites.

From the selected independent variables, an empirical approach was employed to identify the most appropriate curve that fitted the data. Linear, cubic, quadratic and power curve functions were fitted to the data. The R^2, Adj-R^2 and RMSE were used to determine the most appropriate model. Of the four non-linear functions tested, the power function was found to best fit the sample data (Table 5).

ANCOVA for pooling data from different forest sites

The regression output of equation (1) fit to the 209 individual trees from both sites and the F

Table 5 Comparison of the curve fitted transformation in nonlinear regression model from pooled sample data

Independent variable	Model	R^2	Adj-R^2	RMSE (kg tree^{-1})
CW	Linear	0.385	0.382	1745.172
	Cubic	0.395	0.386	1739.572
	Quadratic	0.386	0.380	1748.366
	Power	0.450	0.447	1.161
h80	Linear	0.245	0.241	1933.415
	Cubic	0.254	0.244	1930.686
	Quadratic	0.251	0.244	1929.913
	Power	0.532	0.530	1.071

CW = crown width, h = percentiles, R^2 = coefficient of determination, Adj-R^2 = adjusted coefficient of determination, RMSE = root -mean-square error

test for the overall model was statistically highly significant ($p \leq 0.001$) and provided a good fit to the data (Table 6). It was found that all the interaction effects, Ln (CW) * dummy site and Ln (h80) * dummy site were not statistically significant according to their p values (0.782 and 0.339 respectively), which supported the null hypothesis: the response variables of the AGB in PFR and FRIM were independent of the site factor. Therefore, there were no statistically significant changes in the regression slopes of the model in the presence of different data from FRIM site. This result confirmed an early assumption that the vegetation structure was likely to have more impact on the LiDAR metrics than site-level environmental factors. Therefore, the data from FRIM could be pooled with the PFR data to develop the regression models and to mitigate the issue of the limited sample size in this study.

Data outlier and influential analysis

The results from the regression are shown in Table 7. The p-value and R^2 value for the model were ≤ 0.001 and 0.6112, respectively, thus the model provided a good fit to the data.

Regression model fitting analysis

Multicollinearity is a serious issue that must be considered when using regression models. Multicollinearity was controlled by checking the VIF of the models. All independent variables and the model as a whole had a VIF below 5. Normal probability plots of the residuals suggested that the residuals were normally distributed. Besides the normal Q-Q plot distribution, residuals from the prospective model were also tested for normality by using the Shapiro-Wilk test and analysis on heteroscedasticity by using Breusch-Pagan test. Both tests accept the null hypothesis that residuals from the model are normally distributed by reporting $p = 0.13$ for Shapiro-Wilk test, and the variance of the error term is constant (homoscedasticity) with $p = 0.34$.

By adopting the power transformation into the model's design, the residual variances were reduced and the nonlinearity was addressed. Further, the linearity was not violated according to the normal probability plot of the residuals and the power function model was best fit to the sample data. Next, how well the proposed model could predict the outcome in different datasets (known as model generalisability) need to be assessed. The generalisation of the proposed regression model was evaluated using the 95% confidence intervals of the prediction in addition to cross-validation. The model performed well for small predicted values (Figure 6).

The mean square error (MSE) from a ten-fold cross validation was 0.907 compared to 0.856 that was estimated from the 105 samples. Since the two results were reasonably close, the proposed model can be considered generalisable. Figure 7, shows the model's predictive results for both the training and calibration data sets in each pass. Prediction of the omitted test data are used to assess the predictive accuracy. The dashed lines are parallel and close to each other. The model's curve were scattered gradually in the low AGB values and close to each other towards the high AGB range, slightly showing a positive trend of errors in the low AGB range.

Table 6 Result of analysis of covariance derived from 209 individual trees

Source	Sum of squared (type III)	Df	F value	Significance
Intercept	0.044	1	0.049	0.8250
Ln (CW)	39.546	1	43.861	3.04e-10
Ln (h80)	71.087	1	78.844	3.41e-16
Ln (CW) * dummy site	0.069	1	0.077	0.782
Ln (h80) * dummy site	0.827	1	0.917	0.339

*Residual standard error = 0.9495 on 205 degrees of freedom, Adj-R^2 = 0.630, F-statistic = 90.04 on 4 and 205 DF, p-value < 2.2e-16, CW = crown width, h = percentiles, Ln = natural logarithm

Table 7 Results of nonlinear power model

	Coefficients	Estimated standard error	t value	Pr (> |t|)
(Intercept)	0.612	0.463	1.324	0.187
Ln (CW)	0.629	0.085	7.375	3.9e-12
Ln (h80)	1.540	0.154	10.012	<2e-16

Ln = natural logarithm

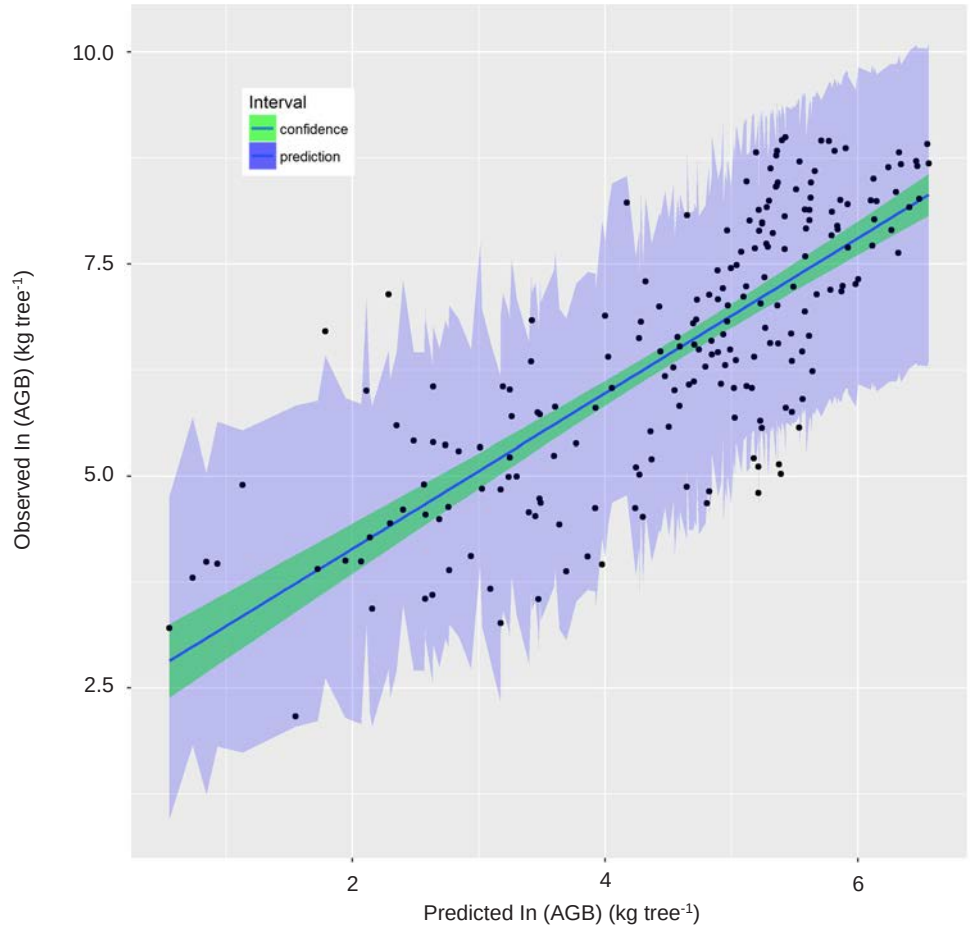


Figure 6 The 95% confidence intervals of the predictions of power regression model, 95% probability that the true best-fit line for aboveground biomass (AGB) lies within the confidence interval and 95% of the y-values to be found for a certain x-value within the interval range around the linear regression line; ln = natural logarithm

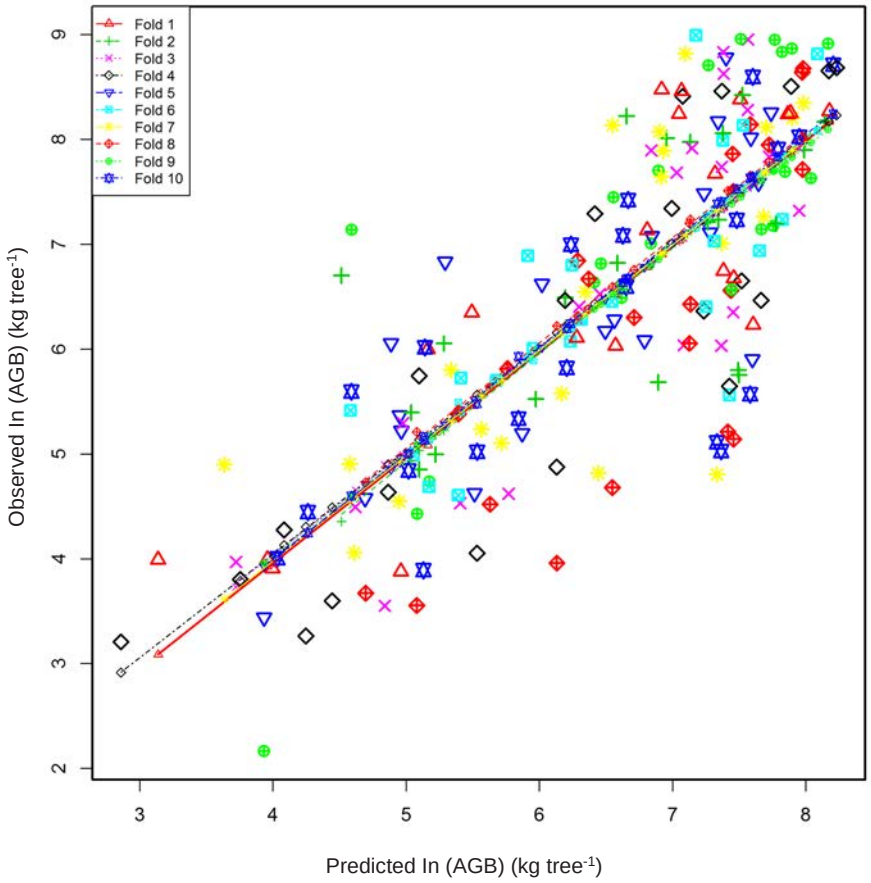


Figure 7 Summary of 10-fold cross validation for evaluating aboveground biomass (AGB) power model, the line is fitted to the training data set in each pass, leaving out corresponding test data set; ln = natural logarithm

AGB from at the two sites was well estimated using the final prediction model predicting AGB from two LiDAR metrics (i.e. CW and h80), with Adj-R^2 of 0.63 and RMSE of 14.68% (Figure 8).

DISCUSSION

Malaysia's interest in participating in the World Bank's Forest Carbon Partnership Facility (FCPF) (Hamdan 2012) requires a baseline study to calculate carbon, so as to present the country's current status of carbon stored in forest. The development of biomass model for carbon stock estimation will help to support the research development for carbon monitoring methodology in Malaysia, with aims to prepare forested countries for REDD+ implementation in Malaysia (FRIM 2011, FRIM 2012). To best of our knowledge, this is one of the first studies to develop a model using individual tree and LiDAR-derived crown metrics at tree level in tropical rainforest. The increasing importance of accurate biomass estimation to support the REDD+ implementation, has created a critical need to understand, evaluate and improve current tree biomass prediction methods by adopting state-of-the-art analytical and statistical techniques. In this study, LiDAR metrics that correlate very well with field AGB was extracted based on trees that were mapped in the field as the main point to develop the LiDAR-AGB model.

Field-LiDAR AGB model

Field sample data can be valuable for evaluating LiDAR-based and other remotely sensed AGB maps, as plots are systematically arranged to provide a spatially unbiased estimate of forest AGB over an area, followed by well-documented measurement protocols that are quality controlled (Johnson et al. 2014). Findings from this study show that the LiDAR metrics and stem-localised AGB correlated very well

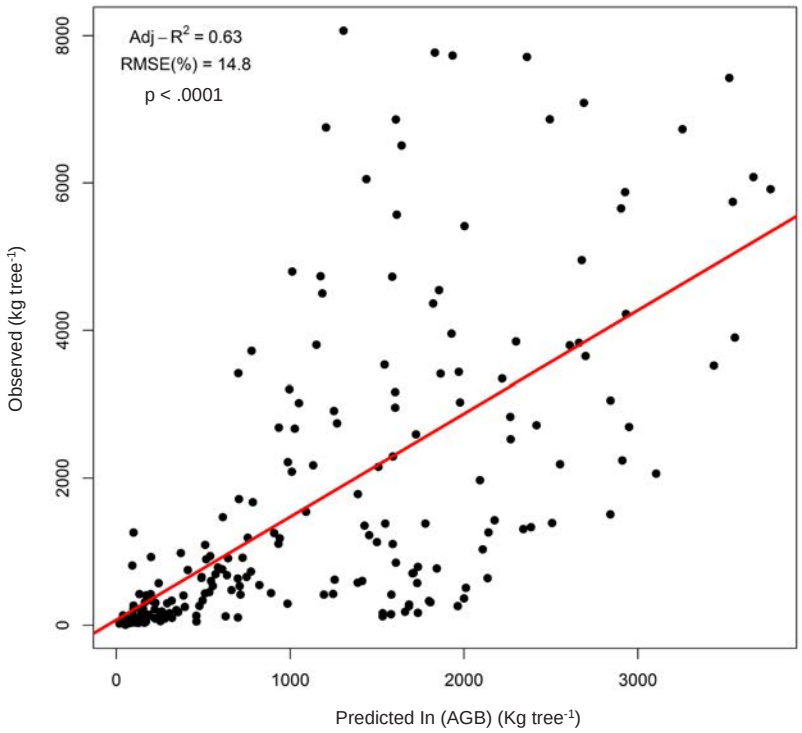


Figure 8 Scatter plot of final observed *versus* predicted aboveground biomass (AGB) from the combination at both sites

with < 15% error in the residuals. An interesting finding from this research was that CW was one of the best LiDAR metrics for predicting AGB. Incorporating both crown size and tree height, particularly of large trees, may improve estimates from remote sensing data for both standing carbon stocks and carbon stock changes and this may be especially applicable to methods based on small footprint LiDAR (Goodman et al. 2014). A recent published study by Ferraz et al. (2016) also incorporates CW, from decomposed entire point cloud into 3-Dimensional (3D) clusters that correspond to individual tree crown. The 3D clusters were modelled using a convex hull to calculate the crown area (CA) and the crown volume (CV) using the same tools developed by Silva et al. (2015), as used in this study. However, the method allows for the use of many forest parameters in existing field allometrics equations to estimate the AGB. This method is only appropriate for forest areas with well-established field AGB allometry. The model developed was based on a geolocated stem map that represent most of the species which exist for lowland dipterocarp forest for building up LiDAR-AGB model with an adequate comparison of individual tree at stem level. Developing a regression model for AGB estimation at stem level is important to derive the most important LiDAR metrics that correlate with field AGB, before implementation at landscape level. The initial approach for quality assessment was presented in this study.

Herein CW is related to DBH, direct retrieval of DBH from LiDAR point cloud at tree level is possible at temperate forest but very little has been addressed (Maltamo et al. 2009, Vauhkonen et al. 2010). However, it is not possible to apply the approach to tropical rainforest. Instead, an interpolation through the relationship with crown size is needed. Given that extraction of CW can be achieved through LiDAR, it is possible that future studies could retrieve DBH using similar methods. The CW and DBH relationships of tropical forest has previously been reported (Perez 1970), the CW and DBH relationships has also been reported (Wile 1964, Roberts & Ross 1965, Bonner 1968) for a number of conifer species. All of these studies found a strong relationship between DBH and CW (R^2 = 0.6–0.9).

Metrics selection and their explanation

The analysis involved all independent variables in the initial search, finding correlations among independent variables, starting the analysis

using the less correlated metrics ($r < 0.7$) and discarding those with least importance until the classification accuracy became stable. The top-ranked metrics were selected based on the lowest AIC value as the first criteria. To make sure the best model with the lowest AIC value is stable from the statistical perspective, multicollinearity was tested using VIF. The best model based on AIC contained CW, CBH, 5th percentile and 75th percentile. However, it also introduced high multicollinearity based on VIF value > 5 and the p-values were not significant for CBH (0.06) and 5th percentile (0.32) variables. There were also strong high correlations among independent variables as shown in Figure 5(a). Regression models with highly correlated independent variables are not stable and hard to interpret from either statistical or biological perspectives (Naesset & Gobakken 2008). The second best model with the lowest AIC value was the model containing CW and 80th percentile height as the independent variables for predicting AGB. This model passed all the diagnostic tests, with all variables having a VIF less than 5, suggesting no serious multicollinearity in the model. In addition, all variables were significant in terms of p-values, and there were no statistically significant changes in the regression slopes of the model in the presence of combining data from FRIM field sample plot.

There have been other studies reporting different height metrics that correlate very well with AGB that varies not just depending on forest type and location, but also on model and data processing procedures (Thapa et al. 2015). In this study, multiplicative models using power functions performed best for predicting AGB from metrics generated from the raw point cloud LiDAR. AGB is usually non-linearly related to remote sensing variables, therefore, nonlinear transformations such as a power function where the response and explanatory variables were log transformed (Hall et al. 2005) reduces the heterogeneity of the regression residual variance. On the plus side, power functions have been used in a large number of studies for AGB estimation, probably because most allometric equations for calculating AGB in the field are power models (Qi 2013). An empirical approach has been implemented in this study to identify the most appropriate curve that fitted the data and power model was found to be the best fit for the sample data, providing justification for the model specification.

Analysis of covariance and regression model

Although environmental and successional differences existed between PFR and FRIM, the regression output of ANCOVA demonstrated that the relationship between AGB and LiDAR metrics was independent of these two sites by confirming the assumption of homogeneous slopes. High AGB is related to the size and structure of trees. The smaller the sample size, the greater the errors will be; limited sample size was overcome by combining the sample data between two sites similar in nature. The 105 samples from PFR were probably too small to represent the tree species and structural diversity exist in tropical rainforest. To overcome the issue of uncertainty in estimating the regression coefficients and intercepts, pooling sample data from similar forest types is a means to overcome uncertainty in AGB estimation. The final selected model, did not violate assumptions of equal variance and normality. However, there was a slight positive trend in the lower predicted AGB errors confirmed in both the variances of residuals and the ten-fold cross-validation analysis. It implies that the model might provide a better AGB estimation with a higher AGB than in the lower AGB.

This study provided direct retrieval of individual metrics (e.g. tree height and crown size) that allow testing of the hypothesis that LiDAR-based AGB models can replace or complement ground-based AGB models. Ferraz et al. (2016) implemented a similar approach through direct retrieval of LiDAR individual tree metrics, but the model assessment was dependent on a ground-based AGB model, which is an approach only appropriate for a forest area with well-established field AGB allometry. In contrast, the current approach used a precisely geolocated stem map that represented the majority of tropical rainforest species to develop the AGB-LiDAR model. It is much more reliable as it showed that important LiDAR metrics correlated very well with field-based AGB measurements, before implementing at a landscape level.

CONCLUSIONS

ANCOVA analysis confirmed that the relationship between AGB and LiDAR were independent of the forest site, supporting an early assumption that the characteristics of the vegetation structure were likely to have more dominant influence on

LiDAR metrics than environmental and other factors between the study areas. Combining data from two forest sites has two benefits. First tropical forest is a very complicated ecological system with diverse species and different biophysical structures, therefore the 105 tree sampled at PFR were possibly too small to build a reliable regression model to represent the whole population that exist in the region. Since PFR and FRIM forest is similar in terms of topographic conditions, vegetation cover and land use history, it is useful to combine the datasets for model calibration and validation. Second, in terms of statistics, combining data is a great way to understand the result of applying multiple models and approaches. The regression model in this study indicated a good correlation between LiDAR predictors and AGB, as in many prior studies.

A power function was identified as the best solution to fit the modelled data. Of 30 LiDAR metrics, h80 and CW were identified as the best independent variables to predict AGB. An interesting finding from this study, supporting earlier findings of Goodman et al. (2014), was the importance of incorporating crown size and height to improve estimates of AGB carbon in tropical forest especially, also based on small footprint LiDAR. This research provided an analytic framework for developing a predictive AGB model from LiDAR and field plot data and will be of value especially for forest resource managers for estimating the AGB in lowland dipterocarp forest to improve management decisions.

It should stress that, the method used to extract LiDAR data for the trees was not an automatic individual tree detection method, but based on trees that were mapped in the field. The results of this study, may have been affected by the manual tree delineation, and individual tree position on the field was recorded by handheld GPS, however this approach tends to give an over-promising impression of the methods. It is suggested that an automatic procedure with focus to derive multi-layered crown delineation will be a promising avenue for future research. The usefulness of producing accurate tree-level data by means of LiDAR should therefore be assessed carefully with respect to alternative methods, model improvement and the costs involved.

ACKNOWLEDGEMENTS

The authors would like to thank Airborne Informatics Malaysia, vendor of LIDAR data, and Geoinformation Programme team, FRIM, for providing access to Pasoh Forest Reserve, field assistance in data collection and fieldwork instruments and diligence with plot geolocation setup. The authors would also like to thank the survey team from the Faculty of Architecture, Planning and Surveying, MARA University of Technology, Malaysia for setting up GPS plot location during fieldwork campaign. Utmost thanks to Majlis Amanah Rakyat (MARA), Malaysia (330408224838) for supporting the PHD research, and Lamb Fund, Geography Centenary, School of Geosciences, University of Edinburgh, UK (E08802) to have supported the research during fieldwork campaign.

REFERENCES

Anaya JA, Chuvieco E & Palacios-Orueta A. 2009. Aboveground biomass assessment in Colombia: A remote sensing approach. *Forest Ecology and Management* 257: 1237–1246.

Asner GP. 2009. Tropical forest carbon assessment: integrating satellite and airborne mapping approaches. *Environmental Research Letters* 3: 1748–9326

Basuki TM, Van Laake PE, Skidmore AK & Hussin YA. 2009. Allometric equations for estimating the above-ground biomass in tropical lowland Dipterocarp forests. *Forest Ecology and Management* 257: 1684–1694.

Bonner GM. 1968. Stem diameter estimates from crown width and tree height. *Commonwealth Forestry Review* 47: 8–13.

Breusch TS & Pagan AR. 1979. A Simple test for heteroscedasticity and random coefficient variation. *Econometrica* 47: 1287–1294.

Butler RA. 2013. Malaysia has the world's highest deforestation rate, reveals Google forest map. https://news.mongabay.com/2013/11/malaysia-has-the-worlds-highest-deforestation-rate-reveals-google-forest-map/.

Chave J, Rejou-Mechain M, Burquez A et al. 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology* 20: 3177–3190.

Dawkins HC. 1963. Crown diameters: their relation to bole diameter in tropical forest trees. *The Commonwealth Forestry Review* 42: 318–333.

Dubayah RO & Drake JB. 2000. LiDAR remote sensing for forestry. *Journal of Forestry* 98: 44–46.

Ferraz A, Saatchi S, Mallet C & Meyer V. 2016. Lidar detection of individual tree size in tropical forests. *Remote Sensing of Environment* 183: 318–333.

FRIM. 2011. *Research on Development of Caron Monitoring Methodology for REDD+ in Malaysia.* Forest Research Institute Malaysia, Kepong.

FRIM. 2012. *Research on Development of Caron Monitoring Methodology for REDD+ in Malaysia.* Forest Research Institute Malaysia, Kepong.

Gibbs HK, Brown S, Niles JO & Foley JA. 2007. Monitoring and estimating tropical forest carbon stocks: making REDD a reality. *Environmental Research Letters* 2: 045023.

Goodman RC, Phillips OL & Baker TR. 2014. The importance of crown dimensions to improve tropical tree biomass estimates. *Ecological Applications* 24: 680–698.

Hall SA, Burke IC, Box DO, Kaufmann MR & Stoker JM. 2005. Estimating stand structure using discrete-return LiDAR: an example from low density, fire prone ponderosa pine forests. *Forest Ecology and Management* 208*:* 189–209.

Hamdan O. 2012. *An Overview of REDD+ Activities in Malaysia.* Forest Research Institute Malaysia, Kepong.

Hamdan O, Mohd-Hasmadi I, Khali-Aziz H, Norizah K & Helmi-Zulhaidi MS. 2015. Determining L-band saturation level for aboveground biomass assessment of dipterocarp forests in Peninsular Malaysia. *Journal of Tropical Forest Science* 27: 1–12.

Houghton RA, Lawrence KT, Hackler JL & Brown S. 2001. The spatial distribution of forest biomass in the Brazilian Amazon: a comparison of estimates. *Global Change Biology* 7: 731–746.

Johnson KD, Birdsey R, Finley AO et al. 2014. Integrating forest inventory and analysis data into a LiDAR-based carbon monitoring system. *Carbon Balance and Management* 9:1–11.

Johnson KD, Birdsey R, Finley AO et al. 2014. Integrating forest inventory and analysis data into a LiDAR-based carbon monitoring system. *Carbon Balance and Management* 9:1–11.

Kaartinen H, Hyyppä J, Yu X et al. 2012. An international comparison of individual tree detection and extraction using airborne laser scanning. *Remote Sensing* 4: 950–974.

Khosravipour A, Skidmore AK, Isenburg M, Wang TJ & Hussin YA. 2014. Generating pit-free canopy height models from airborne LiDAR. *Photogrammetric Engineering and Remote Sensing* 80: 863–872.

Kraus K & Pfeifer N. 2001. Advanced DTM generation from LiDAR data. *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences* 34: 23–30.

Lefsky MA, Cohen WB, Harding DJ, Parker GG, Acker SA & Gower ST. 2002. Lidar remote sensing of above-ground biomass in three biomes. *Global Ecology and Biogeography* 11: 393–399.

Lefsky MA, Hudak AT, Cohen WB & Acker SA. 2005. Patterns of covariance between forest stand and canopy structure in the Pacific Northwest. *Remote Sensing of Environment* 95: 517–531.

Li WK, Guo QH, Jakubowski MK & Kelly M. 2012. A new method for segmenting individual trees from the LiDAR point cloud. *Photogrammetric Engineering and Remote Sensing* 78: 75–84.

Li YZ, Andersen HE & Mcgaughey R. 2008. A comparison of statistical methods for estimating forest biomass from Light Detection and Ranging data. *Western Journal of Applied Forestry* 23: 223–231.

Lim K, Treitz P, Wulder M, St-Onge B & Flood M. 2003. LiDAR remote sensing of forest structure. *Progress in Physical Geography* 27: 88–106.

Lu DS. 2006. The potential and challenge of remote sensing-based biomass estimation. *International Journal of Remote Sensing* 27: 1297–1328.

Lu D, Chen Q, Wang G et al. 2012. Aboveground forest biomass estimation with landsat and LiDAR data and uncertainty analysis of the estimates. *International Journal of Forestry Research* 2012: 1–16.

Mallet C & Bretar F. 2009. Full-waveform topographic LiDAR: state-of-the-art. *Isprs Journal of Photogrammetry and Remote Sensing* 64: 1–16.

Maltamo M, Peuhkurinen J, Malinen J, Vauhkonen J, Packalén P & Tokola T. 2009. Predicting tree attributes and quality characteristics of Scots pine using airborne laser scanning data. *Silva Fennica* 43: 507–521

Mcgaughey RJ. 2014. Fusion/LDV: software for LiDAR data analysis and visualization version 3.50. http://forsys.cfr.washington.edu/fusion/FUSION_manual.pdf.

Morel AC, Saatchi SS, Malhi Y et al. 2011. Estimating aboveground biomass in forest and oil palm plantation in Sabah, Malaysian Borneo using ALOS PALSAR data. *Forest Ecology and Management* 262: 1786–1798.

Naesset E. 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment* 80: 88–99.

Naesset E, Bollandsas OM & Gobakken T. 2005. Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data. *Remote Sensing of Environment* 94: 541–553.

Naesset E & Gobakken T. 2008. Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sensing of Environment* 112: 3079–3090.

Niiyama K, Kajimoto T, Matsuura Y et al. 2010. Estimation of root biomass based on excavation of individual root systems in a primary dipterocarp forest in Pasoh Forest Reserve, Peninsular Malaysia. *Journal of Tropical Ecology* 26: 271–284.

Niklas KJ & Enquist BJ. 2001. Invariant scaling relationships for interspecific plant biomass production rates and body size. *Proceedings of the National Academy of Sciences of the United States of America* 98: 2922–2927.

Perez JW. 1970. Relation of crown diameter to stem diameter in forest of Puerto Rico, Dominic and Thailand. Pp B105–B122 in Odum HT & Pigeon RF (eds) *A Tropical Rainforest – A Study of Irradiation and Ecology at El Verde.* U.S. Atomic Energy Commission, Division of Technical Information, Springfield.

Qi C. 2013. LiDAR Remote Sensing of Vegetation Biomass. *Remote Sensing of Natural Resources.* CRC Press and Taylor & Francis Group, Boca Raton.

Reutebuch SE, Andersen HE & Mcgaughey RJ. 2005. Light Detection and Ranging (LiDAR): an emerging tool for multiple resource inventory. *Journal of Forestry* 103: 286–292.

Shapiro SS. & Wilk MB. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52: 591–611.

Silva CA, Crookston NL, Hudak AT & Vierling LA. 2015. LiDAR: An R package for reading, processing and visualizing LiDAR (Light Detection and Ranging) data, version 0.1. https://cran.rproject.org/web/packages/rLiDAR/index.html.

Steininger MK. 2000. Satellite estimation of tropical secondary forest above-ground biomass: data from Brazil and Bolivia. *International Journal of Remote Sensing* 21: 1139–1157.

Symington CF. 1943. *Foresters' Manual of Dipterocarps*. University of Malaya, Kuala Lumpur.

Thapa RB, Watanabe M, Motohka T, Shiraishi T & Shimada M. 2015. Calibration of aboveground forest carbon stock models for major tropical forests in central Sumatra using airborne LiDAR and field measurement data. *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing* 8: 661–673.

UNFCC 2007. Reducing emissions form deforestation in developing countries: approaches to stimulate action, conference of the parties. Paper presented at The United Nations Framework Convention in Climate Change Conference, Bonn.

UNFCCC 2009. Recommendations on future financing options for enhancing the development, deployment, diffusion and transfer of technologies under the convention. Paper presented at The United Nations Framework Convention in Climate Change Conference, Bonn.

Vauhkonen J, Korpela I, Maltamo M & Tokola T. 2010. Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. *Remote Sensing of Environment* 114: 1263–1276.

William FL. 2007. Forest destruction in tropical Asia. *Current Science* 93: 1544–1550.

Wang X, Ouyang S, Sun OJ & Fang J. 2013. Forest biomass patterns across northeast China are strongly shaped by forest height. *Forest Ecology and Management* 293: 149–160.

Wile BC. 1964. *Crown Size and Stem Diameter in Red Spruce and Balsam Fir*. Department of Forestry Publication, Ottawa.

Wyatt-Smith J. 1964. A preliminary vegetation map of Malaya with descriptions of the vegetation types. *Journal of Tropical Geography* 18: 200–213.

APPENDIX 2

Pasoh Forest Reserve Field Data

Tree No	Species Name	Family	density (g/cm3)	DBH (cm)	Height (m)	AGB (kg) Chave et al. (2014)	AGB (kg) Kato et al. (1978)	AGB (kg) Basuki et al. (2009)
1	Madhuca pasohensis	Sapotaceae	0.57	17.1	15.9	147.61	138.38	153.49
2	Shorea multiflora	Dipterocarpaceae	0.52	35.7	18.5	658.27	683.62	772.81
3	Xanthophyllum stipitatum	Polygalaceae	0.84	12.1	12.9	89.46	57.33	71.81
4	Archidendron bubalinum	Fabaceae	0.38	38.7	21.5	656.99	929.31	922.63
5	Teijsmanniodendron coriaceum	Lamiaceae	0.5	12.4	11.3	49.70	52.87	75.78
6	Shorea macroptera	Dipterocarpaceae	0.39	26.9	21.9	337.32	462.24	415.10
8	Madhuca pasohensis	Sapotaceae	0.57	12.7	9.3	48.93	45.86	79.87
9	Ctenolophon parvifolius	Ctenolophonaceae	0.72	26.2	16.2	434.29	326.16	391.74
11	Diospyros kaki	Ebenaceae	0.551	39.4	21.6	982.23	967.14	959.67
12	Knema kunstleri	Myristicaceae	0.57	11.6	7.9	34.97	32.95	65.46

13	Neoscortechinia kingii	Euphorbiaceae	0.53	14.2	9.5	57.87	58.13	102.06
15	Dipterocarpus sublamellatus	Dipterocarpaceae	0.63	82.3	27.9	6052.60	5320.51	4837.60
16	Alangium ebenaceum	Alangiaceae	0.57	13.2	8.5	48.33	45.30	86.93
17	Litsea ferruginea	Lauraceae	0.51	15.4	15	101.98	106.42	121.96
18	Syzygium duthianeum	Myrtaceae	0.74	14.7	16.5	146.98	106.65	110.11
19	Sindora echinocalyx	Fabaceae	0.61	54.1	31.5	2911.05	2621.16	1925.37
20	Monocarpia marginalis	Annonaceae	0.48	17.6	13.7	114.18	126.52	163.52
21	Reinwardteodendron cinereum	Meliaceae	0.57	13.7	8.9	54.35	50.89	94.33
22	Teijsmanniodendron coriaceum	Lamiaceae	0.79	16.3	11.5	134.75	91.69	138.16
23	Monocarpia marginalis	Annonaceae	0.48	22.1	11.6	151.38	168.03	269.59
24	Shorea macroptera	Dipterocarpaceae	0.39	14.8	19.6	94.29	127.96	111.77
25	Xantophyllum stipitatum	Polygalaceae	0.84	33.7	18.5	939.30	610.22	680.91
26	Polyalthia hypoleuca	Annonaceae	0.64	14.8	12.9	101.66	84.93	111.77
27	Monocarpia marginalis	Annonaceae	0.48	10.9	7.9	26.18	29.28	57.10
28	Santiria laevigata	Burseraceae	0.56	19.1	14.5	164.56	157.07	195.69
29	Neoscortechinia kingii	Euphorbiaceae	0.53	16.8	12.9	108.30	108.86	147.64

31	Blumeodendron tokbrai	Euphorbiaceae	0.53	29.6	16	403.73	409.67	512.12
32	Uvaria foetida	Annonaceae	0.57	19.4	15.2	180.72	169.64	202.50
33	Polyalthia jenkinsii	Annonaceae	0.57	14.5	12.6	85.25	79.74	106.85
34	Dacryodes costata	Burseraceae	0.73	10.7	11.5	54.86	40.44	54.82
35	Dipterocarpus sublamellatus	Dipterocarpaceae	0.63	81.8	31.6	6753.99	5944.22	4773.30
36	Archidendron microcarpum	Leguminosae	0.57	19.2	15.9	185.06	173.73	197.94
37	Cyathocalyx pruniferus	Annonaceae	0.39	29.3	29.5	533.05	733.50	500.79
38	Hopea mengarawan	Dipterocarpaceae	0.57	12.2	11.9	57.55	53.87	73.12
40	Drypetes kikir	Euphorbiaceae	0.72	18.1	16.7	217.34	162.36	173.89
42	Dillenia sumatrana	Dilleniaceae	0.68	13.9	11.3	83.85	66.02	97.38
43	Shorea multiflora	Dipterocarpaceae	0.52	71.4	27.9	3803.32	4019.84	3541.06
44	Syzygium gratum	Myrtaceae	0.57	28.7	15.9	405.60	383.14	478.54
45	Cleidion spiciflorum	Euphorbiaceae	0.71	13.5	12.6	91.88	69.35	91.33
46	Dacryodes costata	Burseraceae	0.73	28.6	21.3	682.25	507.45	474.89
47	Aporosa prainiana	Phyllanthaceae	0.62	17.3	9.3	97.12	83.69	157.46
48	Callerya atropurpurea	Fabaceae	0.56	16.4	15.2	127.97	121.97	140.03

49	Chisocheton erythrocarpus	Meliaceae	0.57	22.2	14.5	224.56	211.10	272.27
50	Chisocheton erythrocarpus	Meliaceae	0.57	10.3	8.9	31.15	29.44	50.42
51	Anacolosa frutescens	Olacaceae	0.57	14.1	11.2	71.96	67.30	100.48
52	Myristica cinnamomea	Myristicaceae	0.57	12.4	8.9	44.74	41.97	75.78
53	Aporosa bracteosa	Phyllanthaceae	0.57	10.9	10.1	39.36	36.99	57.10
54	Xantophyllum rufum	Polygalaceae	0.57	38.9	16.5	761.37	723.30	933.13
55	Dacryodes costata	Burseraceae	0.73	11.1	10.2	52.42	38.67	59.42
56	Teijsmanniodendron coriaceum	Lamiaceae	0.79	13.5	9.2	75.02	51.08	91.33
57	Knema scortechinii	Myristicaceae	0.57	14.7	14.9	103.13	96.51	110.11
58	Dacryodes costata	Burseraceae	0.73	30.4	20.8	750.96	559.05	543.00
59	Cryptocarya rugulosa	Lauraceae	0.57	10.6	9.9	36.55	34.41	53.70
61	Garcinia eugeniifolia	Guttiferae	0.57	10.5	9.6	34.82	32.81	52.59
62	Gironniera parvifolia	Ulmaceae	0.57	11.6	13.9	60.70	56.79	65.46
63	Shorea leprosula	Dipterocarpaceae	0.47	41.1	26.3	1106.82	1276.17	1052.95
64	Santiria apiculata	Burseraceae	0.53	17.5	14.8	134.12	134.96	161.48
65	Dipterocarpus cornutus	Dipterocarpaceae	0.66	72.7	26.9	4797.81	4018.20	3684.19

66	Syzygium duthianeum	Myrtaceae	0.74	14.9	13.8	126.76	91.93	113.43
67	Shorea acuminata	Dipterocarpaceae	0.48	40.6	21.3	897.94	1012.01	1025.03
68	Shorea leprosula	Dipterocarpaceae	0.42	53.4	29.5	1849.32	2394.67	1871.08
69	Monocarpia marginalis	Annonaceae	0.48	34.8	19.3	603.63	677.76	730.67
70	Dipterocarpus cornutus	Dipterocarpaceae	0.66	13.4	8	54.13	43.98	89.85
71	Nephelium hamulanthum	Sapindaceae	0.57	10.1	7.3	24.71	23.58	48.29
73	Reinwardteodendron cinereum	Meliaceae	0.57	16.1	15.9	131.23	122.94	134.46
74	Dipterocarpus cornutus	Dipterocarpaceae	0.66	77.9	39.9	8067.16	6794.73	4287.74
75	Monocarpia marginalis	Annonaceae	0.48	23.6	20.1	294.27	328.31	311.41
76	Xylopia magna	Annonaceae	0.64	39.1	23	1190.75	1013.50	943.70
77	Callerya atropurpurea	Fabaceae	0.56	12.1	11.3	52.92	50.42	71.81
78	Koompassia malaccensis	Fabaceae	0.76	109.9	32.8	14969.58	11046.41	9129.42
79	Diospyros buxifolia	Ebenaceae	0.72	13.7	14.5	109.94	81.87	94.33
83	Polyalthia Hypoleuca	Annonaceae	0.64	13.1	15	92.82	77.53	85.49
86	Shorea Macroptera	Dipterocarpaceae	0.39	28.3	19.6	334.21	457.94	464.02
89	Santiria Laevigata	Burseraceae	0.56	22	16.8	250.36	239.69	266.91

94	Koompassia Malaccensis	Leguminosae	0.76	71.3	33.2	6509.55	4759.09	3530.18
95	Shorea Macroptera	Dipterocarpaceae	0.39	68.9	31.5	3016.12	4223.41	3274.48
102	Knema Kunstleri	Myristicaceae	0.57	19.5	17.9	214.13	201.23	204.80
106	Shorea Leprosula	Dipterocarpaceae	0.42	53.7	26.3	1671.45	2162.09	1894.24
111	Diospyros Boxifolia	Ebenaceae	0.72	20.6	17.9	299.38	224.17	231.03
114	Shorea Leprosula	Dipterocarpaceae	0.42	60.9	27.3	2216.02	2874.90	2497.08
115	Santiria Apiculata	Burseraceae	0.53	39.5	20.7	911.70	932.07	965.03
116	Pentaspadon Motleyi	Anacardiaceae	0.5	60.3	22.7	2152.14	2350.26	2443.37
117	Ctenolophon Parvifolius	Ctenolophonaceae	0.72	28.7	21.9	696.35	525.13	478.54
118	Shorea Leprosula	Dipterocarpaceae	0.42	82.3	30.9	4501.59	5884.62	4837.60
119	Ochanostachys Amentacea	Olacaceae	0.57	56.3	28.5	2670.98	2569.05	2101.50
120	Dipterocarpus Cornutus	Dipterocarpaceae	0.66	46.2	28.4	2087.90	1733.58	1361.34
121	Callerya Atropurpurea	Fabaceae	0.56	41.4	29.1	1470.25	1430.35	1069.90
122	Dacryodes Laxa	Burseraceae	0.68	26.1	16.3	410.13	325.68	388.47
123	Anicosanthum Fuscum	Annonaceae	0.57	23.1	16.1	268.77	253.02	297.10
124	Artocarpus Elasticus	Moraceae	0.35	36.1	17.2	425.75	650.39	791.95

125	Monocarpia Marginalis	Annonaceae	0.48	29.4	20	449.70	503.57	504.55
126	Mesua Ferrea	Guttiferae	0.9	22.7	16.9	425.34	256.42	285.92
127	Anisophyllea	Anisophylleaceae	0.57	19.4	16.8	199.27	187.17	202.50
128	Ochanostachys Amentacea	Olacaceae	0.57	51.2	21.9	1716.00	1643.08	1705.96
129	Dyera Costulata	Apocynaceae	0.35	43.7	20.4	730.20	1121.21	1204.79
130	Santiria Apiculata	Burseraceae	0.53	26.5	18.5	374.84	380.11	401.66
131	Dipterocarpus Cornutus	Dipterocarpaceae	0.66	27.5	21.3	572.75	469.73	435.70
132	Vatica Bella	Dipterocarpaceae	0.57	19.9	21.9	271.26	255.38	214.14
133	Shorea Acuminata	Dipterocarpaceae	0.48	29.3	21.5	479.39	537.13	500.79
134	Dipterocarpus Cornutus	Dipterocarpaceae	0.66	53.4	27.5	2684.35	2234.51	1871.08
135	Dipterocarpus Sublamellatus	Dipterocarpaceae	0.63	71.3	28.9	4734.11	4150.44	3530.18
136	Ctenolophon Parvifolius	Ctenolophonaceae	0.72	35.6	18.9	918.39	694.33	768.06
137	Cryptocarya Rugulosa	Lauraceae	0.57	26.3	19.5	417.42	394.40	395.03
138	Uvaria Foetida	Annonaceae	0.57	41.1	21.4	1092.58	1041.51	1052.95
139	Trigoniastrum Hypoleucum	Trigoniaceae	0.64	28.2	19.9	546.28	461.61	460.43
140	Canarium Littorale	Burseraceae	0.45	15.3	16.9	100.11	118.10	120.22

141	Hopea Mengarawan	Dipterocarpaceae	0.57	30.1	22.9	635.49	602.71	531.31
142	Dacryodes Costata	Burseraceae	0.73	28.7	19.9	642.81	477.87	478.54
Total AGB (kg) =						**112298.79**	**105022.43**	**92700.36**
Total Carbon Stock ($MgCha^{-1}$) =						**52780.43**	**49360.54**	**92700.36**

Appendix 3

FRIM Forest Reserve Field Data

Tree ID	Species Name	Family	Wood density (g/cm3)	DBH (cm)	Height (m)	AGB (kg) Chave et al. (2014)	AGB (kg) Kato et al. (1978)	AGB (kg) Basuki et al. (2009)
1105160	Scorodocarpus borneensis	OLACACEAE	0.96	25.8	25	852.28	494.13	378.73
1105295	Scorodocarpus borneensis	OLACACEAE	0.96	45.9	35	3644.02	2119.12	1342.00
1106120	Scorodocarpus borneensis	OLACACEAE	0.96	21.5	12	291.68	169.37	253.77
1106121	Dipterocarpus baudii	DIPTEROCARPACEAE	0.64	65.7	12	1737.86	1499.17	2949.75
1106122	Scorodocarpus borneensis	OLACACEAE	0.96	26.4	17	611.79	354.77	398.34
1106123	Scorodocarpus borneensis	OLACACEAE	0.96	21.6	10	246.36	143.12	256.37
1106124	Scorodocarpus borneensis	OLACACEAE	0.96	40.3	15.6	1284.57	744.86	1008.47
1106125	Scorodocarpus borneensis	OLACACEAE	0.96	13.3	16.5	155.86	90.67	88.39
1106126	shorea macroptera	DIPTEROCARPACEAE	0.39	35.5	16	426.74	595.99	763.33
1106127	Scorodocarpus borneensis	OLACACEAE	0.96	42.5	14	1282.18	743.47	1133.32

1106132	Dipterocarpus baudii	DIPTEROCARPACEAE	0.64	49.3	32	2584.11	2232.84	1570.02
1106133	Scorodocarpus borneensis	OLACACEAE	0.96	10.3	18	103.02	60.03	50.42
1106137	Scorodocarpus borneensis	OLACACEAE	0.96	18.1	16.5	284.42	165.17	173.89
1106138	Scorodocarpus borneensis	OLACACEAE	0.96	36	16	1056.42	612.49	787.14
1106168	Scorodocarpus borneensis	OLACACEAE	0.96	14.4	30	326.22	189.37	105.24
1106170	Scorodocarpus borneensis	OLACACEAE	0.96	18.1	35.5	600.79	348.40	173.89
1106176	Scorodocarpus borneensis	OLACACEAE	0.96	14.7	29	328.56	190.73	110.11
1106178	Scorodocarpus borneensis	OLACACEAE	0.96	16.3	33.6	464.10	269.23	138.16
1106181	Dipterocarpus baudii	DIPTEROCARPACEAE	0.64	48	26.9	2070.46	1787.25	1480.53
1106183	Scorodocarpus borneensis	OLACACEAE	0.96	19.1	34	639.75	370.97	195.69
1106184	Scorodocarpus borneensis	OLACACEAE	0.96	27.5	12	471.60	273.57	435.70
1106186	Dipterocarpus baudii	DIPTEROCARPACEAE	0.64	54.9	39	3866.97	3348.85	1988.44
1106190	Dipterocarpus baudii	DIPTEROCARPACEAE	0.64	44.1	40	2584.65	2233.31	1229.13
1106191	Scorodocarpus borneensis	OLACACEAE	0.96	22.5	39	1007.05	583.86	280.42
1106193	Scorodocarpus borneensis	OLACACEAE	0.96	23.4	45	1250.13	724.88	305.64
1106194	Dipterocarpus baudii	DIPTEROCARPACEAE	0.64	72.1	39	6582.87	5722.66	3617.75

1106196	Scorodocarpus borneensis	OLACACEAE	0.96	32.6	37.5	1998.79	1159.88	633.05
1106197	Scorodocarpus borneensis	OLACACEAE	0.96	27.5	31	1190.85	690.48	435.70
1106199	Dipterocarpus baudii	DIPTEROCARPACEAE	0.64	78.6	30	6030.90	5239.20	4372.81
1106200	Scorodocarpus borneensis	OLACACEAE	0.96	25.5	33.9	1121.38	650.17	369.13
1106202	Scorodocarpus borneensis	OLACACEAE	0.96	36	45	2898.34	1683.84	787.14
1106205	Scorodocarpus borneensis	OLACACEAE	0.96	23	40	1077.49	624.71	294.28
1106218	Scorodocarpus borneensis	OLACACEAE	0.96	37.7	33.2	2357.01	1368.37	871.08
1106225	Dipterocarpus baudii	DIPTEROCARPACEAE	0.64	41.8	37.2	2168.87	1872.55	1092.74
1106227	Dipterocarpus baudii	DIPTEROCARPACEAE	0.64	62.2	43	5427.39	4711.14	2615.63
1106232	Scorodocarpus borneensis	OLACACEAE	0.96	23.4	49	1358.48	787.76	305.64
1106234	Scorodocarpus borneensis	OLACACEAE	0.96	21	47	1055.97	612.23	240.99
1106236	Scorodocarpus borneensis	OLACACEAE	0.96	13.2	46	417.76	242.39	86.93
1106237	Scorodocarpus borneensis	OLACACEAE	0.96	24.4	36.1	1094.02	634.29	335.06
1106238	Dipterocarpus baudii	DIPTEROCARPACEAE	0.64	66.4	38.9	5591.26	4854.47	3019.21
1106239	Scorodocarpus borneensis	OLACACEAE	0.96	42.1	40.5	3549.66	2063.99	1110.03
1106240	Dipterocarpus baudii	DIPTEROCARPACEAE	0.64	52.7	42	3838.07	3323.66	1817.64

1106242	Dipterocarpus baudii	DIPTEROCARPACEAE	0.64	63.4	33.2	4376.79	3793.42	2727.72
1106251	Scorodocarpus borneensis	OLACACEAE	0.96	37.5	31.9	2243.47	1302.27	860.96
1106254	shorea macroptera	DIPTEROCARPACEAE	0.39	52.9	34.9	1990.11	2791.80	1832.82
1106258	Scorodocarpus borneensis	OLACACEAE	0.96	17.2	29	446.44	259.00	155.47
1106261	Scorodocarpus borneensis	OLACACEAE	0.96	17.7	40.5	654.08	379.27	165.56
1106268	Scorodocarpus borneensis	OLACACEAE	0.96	24.6	32.9	1015.31	588.65	341.12
1106270	Scorodocarpus borneensis	OLACACEAE	0.96	40.9	43	3556.81	2068.17	1041.73
1106272	Scorodocarpus borneensis	OLACACEAE	0.96	25	39.5	1252.48	726.24	353.42
1106273	Scorodocarpus borneensis	OLACACEAE	0.96	44.4	48	4648.31	2706.51	1247.57
1106275	Scorodocarpus borneensis	OLACACEAE	0.96	20.8	39.9	883.33	512.12	235.98
1106277	Scorodocarpus borneensis	OLACACEAE	0.96	35	45	2743.27	1593.42	739.92
1106280	Scorodocarpus borneensis	OLACACEAE	0.96	58.7	36.9	6201.70	3617.30	2303.25
1106284	Scorodocarpus borneensis	OLACACEAE	0.96	16	43	569.40	330.22	132.64
1106285	Scorodocarpus borneensis	OLACACEAE	0.96	17.2	24.3	375.68	218.02	155.47
1106286	Scorodocarpus borneensis	OLACACEAE	0.96	24.2	42	1247.99	723.63	329.06
1106288	Scorodocarpus borneensis	OLACACEAE	0.96	12.2	37	289.64	168.18	73.12

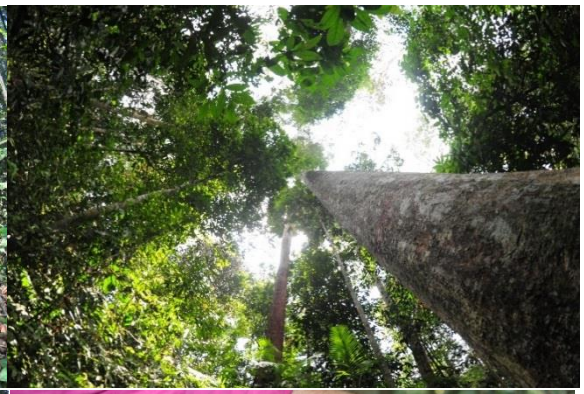
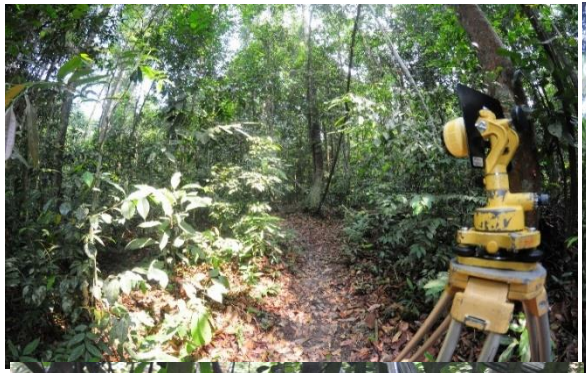
1106291	Dipterocarpus baudii	DIPTEROCARPACEAE	0.64	36.2	38	1672.30	1442.43	796.78
1106292	Timonus wallichianus	RUBIACEAE	0.57	11.9	39.9	178.55	172.44	69.23
1106293	Dipterocarpus baudii	DIPTEROCARPACEAE	0.64	66.4	30	4339.00	3760.44	3019.21
1106294	Actinodaphne pruinosa	LAURACEAE	0.57	16.1	38	307.14	296.31	134.46
1106295	Dipterocarpus baudii	DIPTEROCARPACEAE	0.64	71.3	37	6118.46	5315.86	3530.18
1106296	shorea macroptera	DIPTEROCARPACEAE	0.39	26.7	39.9	597.03	834.04	408.35
1106297	Streblus elongatus	MORACEAE	0.81	11.4	37	214.95	147.38	63.01
1106332	shorea macroptera	DIPTEROCARPACEAE	0.39	64.8	33	2800.00	3936.35	2861.75
1106333	shorea macroptera	DIPTEROCARPACEAE	0.39	63.5	29.8	2436.37	3422.00	2737.18
1106336	Dipterocarpus baudii	DIPTEROCARPACEAE	0.64	28.5	38.2	1053.90	908.01	471.25
1106338	Elateriospermum tapos	EUPHORBIACEAE	0.65	18.6	31.9	390.13	331.04	184.61
1106339	Dipterocarpus baudii	DIPTEROCARPACEAE	0.64	37.8	36.9	1768.19	1525.42	876.16
1107065	Shorea bracteolata	DIPTEROCARPACEAE	0.53	81.9	26	4727.73	4934.88	4786.12
1107066	Shorea bracteolata	DIPTEROCARPACEAE	0.53	63.2	25	2743.45	2853.28	2708.86
1107072	Shorea bracteolata	DIPTEROCARPACEAE	0.53	49.5	33	2232.79	2319.60	1584.04
1107079	Shorea bracteolata	DIPTEROCARPACEAE	0.53	51.5	35	2554.86	2656.07	1727.99

1107082	Shorea bracteolata	DIPTEROCARPACEAE	0.53	78.9	37.9	6349.72	6643.60	4409.54
1107083	Elateriospermum tapos	EUPHORBIACEAE	0.65	37	37.9	1767.26	1501.64	835.96
1107085	Pellacalyx saccardianus	RHIZOPHORACEAE	0.46	29.5	38.1	814.60	968.87	508.33
1107088	Paropsia vareciformis	FLACOURTIACEAE	0.67	35	33.9	1464.74	1207.53	739.92
1107089	Shorea bracteolata	DIPTEROCARPACEAE	0.53	64.3	35.5	3995.39	4165.33	2813.48
1107090	Gomphandra quadrifida	ICACINACEAE	0.53	14	27	156.01	161.79	98.93
1107093	Ochanostachys amentacea	OLACACEAE	0.77	10	34.2	146.70	105.79	47.25
1107103	Shorea bracteolata	DIPTEROCARPACEAE	0.53	69.8	32.9	4354.09	4542.09	3369.14
1107105	Shorea bracteolata	DIPTEROCARPACEAE	0.53	54.2	40.9	3286.37	3421.68	1933.19
1107106	Myristica sp.	MYRISTICACEAE	0.57	37.7	40.9	1737.04	1678.49	871.08
1107114	Shorea bracteolata	DIPTEROCARPACEAE	0.53	64	38	4230.97	4412.74	2784.73
1107118	Shorea bracteolata	DIPTEROCARPACEAE	0.53	41.2	44	2066.32	2145.86	1058.59
1107126	Shorea bracteolata	DIPTEROCARPACEAE	0.53	55.8	40.9	3478.41	3622.94	2060.73
1107129	Scorodocarpus borneensis	OLACACEAE	0.72	21.1	15	264.02	202.92	243.52
1107133	Diospyros wallichii	EBENACEAE	0.6	30.1	35.2	1016.43	932.71	531.31
1107160	Shorea bracteolata	DIPTEROCARPACEAE	0.53	69.3	37.5	4878.36	5093.33	3316.37

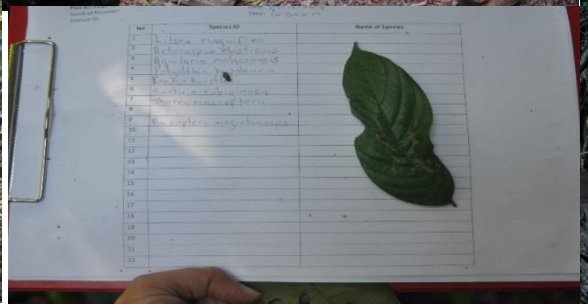
1107167	Barringtonia fusiformis	LECYTHIDACEAE	0.57	13.9	47	283.71	273.74	97.38
1107168	Streblus elongatus	MORACEAE	0.81	17.7	25.9	358.20	245.31	165.56
1107171	Shorea bracteolata	DIPTEROCARPACEAE	0.53	73	39.9	5736.73	5997.19	3717.66
1107193	Shorea bracteolata	DIPTEROCARPACEAE	0.53	67.9	27.5	3463.43	3607.25	3171.02
1107195	Palaquium gutta	SAPOTACEAE	0.57	11.1	20.5	81.38	78.78	59.42
1107208	Artocarpus elasticus	MORACEAE	0.36	20	32	253.28	382.53	216.51
1107214	Archidendron bubalinum	LEGUMINOSAE	0.38	16.7	37	216.37	310.05	145.71
1107215	Shorea bracteolata	DIPTEROCARPACEAE	0.53	82.2	25	4582.76	4782.44	4824.71
1107216	Shorea bracteolata	DIPTEROCARPACEAE	0.53	67.3	36.9	4535.36	4732.60	3109.81
1107222	Shorea bracteolata	DIPTEROCARPACEAE	0.53	60.8	42.9	4308.90	4494.60	2488.08
1107225	Shorea bracteolata	DIPTEROCARPACEAE	0.53	67.5	31.5	3908.95	4074.60	3130.14
1107228	Gomphandra quadrifida	ICACINACEAE	0.57	11	38.9	149.40	144.35	58.25
1107232	Shorea bracteolata	DIPTEROCARPACEAE	0.53	66.8	40.9	4942.13	5160.42	3059.30
1107242	Barringtonia fusiformis	LECYTHIDACEAE	0.57	13.6	22	129.60	125.28	92.82
					Total AGB (Kg) :	215516.94	128215.13	188437.62
					Total Carbon Stock ($MgCha^{-1}$) =	101292.96	60261.11	88565.68

Appendix 4 – Fieldwork photos

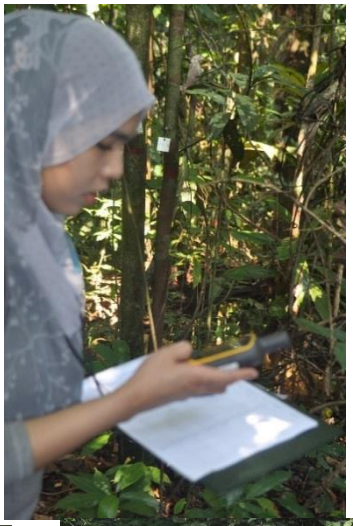










283
15.3
19.6




Stesen Penyelidikan FRIM Pasoh







Stesen Penyelidikan FRIM Pasoh
U

REFERENCES

Achard, F., Eva, H. D., Stibig, H. J., Mayaux, P., Gallego, J., Richards, T., & Malingreau, J. P. (2002). Determination of deforestation rates of the world's humid tropical forests. Science, 297(5583), 999-1002. doi:10.1126/science.1070656

Akaike, H. (1974). A new look at the statistical model identification. IEEE transactions on automatic control, 19(6), 716-723.

Alexander, C. (2009). Delineating tree crowns from airborne laser scanning point cloud data using Delaunay triangulation. International Journal of Remote Sensing, 30(14), 3843-3848.

Amiri, N., Yao, W., Heurich, M., Krzystek, P., & Skidmore, A. K. (2016). Estimation of regeneration coverage in a temperate forest by 3D segmentation using airborne laser scanning data. International Journal of Applied Earth Observation and Geoinformation, 52, 252-262.

Anaya, J. A., Chuvieco, E., & Palacios-Orueta, A. (2009). Aboveground biomass assessment in Colombia: A remote sensing approach. Forest Ecology and Management, 257(4), 1237-1246.

Andersen, H. E., Reutebuch, S. E., & McGaughey, R. J. (2006). A rigorous assessment of tree height measurements obtained using airborne lidar and conventional field methods. Canadian Journal of Remote Sensing, 32(5), 355-366.

Asner, G. P. (2009). Tropical forest carbon assessment: integrating satellite and airborne mapping approaches. Environmental Research Letters, 4(3).

Asner, G. P., & Mascaro, J. (2014). Mapping tropical forest carbon: Calibrating plot estimates to a simple LiDAR metric. Remote Sensing of Environment, 140, 614-624.

Asner, G. P., Hughes, R. F., Varga, T. A., Knapp, D. E., & Kennedy-Bowdoin, T. (2009). Environmental and Biotic Controls Over Aboveground Biomass Throughout a Tropical Rain Forest. Ecosystems, 12(2), 261-278. doi:10.1007/s10021-008-9221-5.

Asner, G. P., Palace, M., Keller, M., Pereira, R., Silva, J. N. M., & Zweede, J. C. (2002). Estimating canopy structure in an Amazon Forest from laser range finder and IKONOS satellite observations. Biotropica, 34(4), 483-492.

Asner, G. P., Powell, G. V. N., Mascaro, J., Knapp, D. E., Clark, J. K., Jacobson, J.,Hughes, R. F. (2010). High-resolution forest carbon stocks and emissions in the Amazon. Proceedings of the National Academy of Sciences of the United States of America, 107(38), 16738-16742. doi:10.1073/pnas.1004875107.

ASPRS, L.C. (2004). ASPRS Guidelines Vertical Accuracy Reporting for Lidar Data. Retrieved 20 Feb 2015 http://www.asprs.org/society/committees/lidar/Downloads/Vertical_Accuracy_Reporting_for_Lidar_Data.pdf

Bai, Y., Walsworth, N., Roddan, B., Hill, D. A., Broersma, K., & Thompson, D. (2005). Quantifying tree cover in the forest–grassland ecotone of British Columbia using crown delineation and pattern detection. *Forest Ecology and Management*, *212*(1-3), 92-100.

Barlow, H. S. (2010). [100 Years of Tropical Forest Research: The Story of the Forest Research Institute Malaysia, Francis S. P. Ng]. Journal of the Malaysian Branch of the Royal Asiatic Society, 83(2 (299)), 124-127.

Baskerville, G. (1972). Use of logarithmic regression in the estimation of plant biomass. Canadian Journal of Forest Research, 2(1), 49-53.

Basuki, T. M., van Laake, P. E., Skidmore, A. K., & Hussin, Y. A. (2009). Allometric equations for estimating the above-ground biomass in tropical lowland Dipterocarp forests. Forest Ecology and Management, 257(8), 1684-1694. doi:10.1016/j.foreco.2009.01.027.

Beucher, S., & Meyer, F. (1992). The morphological approach to segmentation: the watershed transformation. Optical Engineering-New York-Marcel Dekker Incorporated-, 34, 433-433.

Bohlin, J., Wallerman, J., & Fransson, J. E. S. (2012). Forest variable estimation using photogrammetric matching of digital aerial images in combination with a high-resolution DEM. Scandinavian Journal of Forest Research, 27(7), 692-699. doi:10.1080/02827581.2012.686625.

Bollandsås, O. M., Maltamo, M., Gobakken, T., & Næsset, E. (2013). Comparing parametric and non-parametric modelling of diameter distributions on independent data using airborne laser scanning in a boreal conifer forest. Forestry, 86(4), 493-501.

Bonner, G. M. (1968). Stem diameter estimates from crown width and tree height. Commonwealth Forestry Review, 47(1), 8-13.

Borgefors, G., Brandtberg, T., & Walter, F. (1999). Forest parameter extraction from airborne sensors. International Archives of Photogrammetry and Remote Sensing, 32(3; SECT 2W5), 151-158.

Bortolot, Z. J. (2006). Using tree clusters to derive forest properties from small footprint lidar data. Photogrammetric Engineering and Remote Sensing, 72(12), 1389-1397.

Bottcher, H., Eisbrenner, K., Fritz, S., Kindermann, G., Kraxner, F., McCallum, I., &

Obersteiner, M. (2009). An assessment of monitoring requirements and costs of 'Reduced Emissions from Deforestation and Degradation'. Carbon Balance Manag, 4,7. doi:10.1186/1750-0680-4-7.

Boyd, D. S., & Danson, F. M. (2005). Satellite remote sensing of forest resources: three decades of research development. Progress in Physical Geography, 29(1), 1-26. doi:10.1191/0309133305pp432ra.

Brack, C. (1999). Tree crown: Forest measurement and Modelling: Department of Forestry, Australian National University, www.anu.adu.au.

Brandtberg, T., Warner, T. A., Landenberger, R. E., & McGraw, J. B. (2003). Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in North America. Remote Sensing of Environment, 85(3), 290-303. doi:10.1016/S0034-4257(03)00008-7.

Brandtberg, T. (2002). Individual tree-based species classification in high spatial resolution aerial images of forests using fuzzy sets. Fuzzy Sets and Systems, 132(3), 371-387.

Breidenbach, J., & Astrup, R. (2014). The semi-individual tree crown approach Forestry applications of airborne laser scanning (pp. 113-133): Springer.

Breidenbach, J., Naesset, E., Lien, V., Gobakken, T., & Solberg, S. (2010). Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. Remote Sensing of Environment, 114(4), 911-924. doi:10.1016/j.rse.2009.12.004.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. Econometrica, 47(5), 1287-1294. doi:10.2307/1911963.

Butler, R. A. (2013). Malaysia has the world's highest deforestation rate, reveals Google forest map. Retrieved from https://news.mongabay.com/2013/11/malaysia-has-the-worlds-highest-deforestation-rate-reveals-google-forest-map/

Center, I. H. R. (2017). Schematic diagram of an airborne LiDAR system acquisition. Retrieved from http://lidar.ihrc.fiu.edu/aboutlidar.html

Chave, J., Andalo, C., Brown, S., Cairns, M. A., Chambers, J. Q., Eamus, D., Yamakura, T. (2005). Tree allometry and improved estimation of carbon stocks and balance in tropical forests. Oecologia, 145(1), 87-99. doi:10.1007/s00442-005-0100-x.

Chave, J., Condit, R., Lao, S., Caspersen, J. P., Foster, R. B., & Hubbell, S. P. (2003). Spatial and temporal variation of biomass in a tropical forest: results from a

large census plot in Panama. Journal of Ecology, 91(2), 240-252. doi:DOI 10.1046/j.1365-2745.2003.00757.x.

Chave, J., Rejou-Mechain, M., Burquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B., Vieilledent, G. (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. Global Change Biology, 20(10), 3177-3190. doi:10.1111/gcb.12629.

Chen, Q. (2013). Lidar remote sensing of vegetation biomass. Remote sensing of natural resources, 399, 399-420.

Chen, Q., Baldocchi, D., Gong, P., & Kelly, M. (2006). Isolating individual trees in a savanna woodland using small footprint lidar data. Photogrammetric Engineering and Remote Sensing, 72(8), 923-932.

Chen, Q., Baldocchi, D., Gong, P., & Kelly, M. (2006). Isolating individual trees in a savanna woodland using small footprint lidar data. Photogrammetric Engineering and Remote Sensing, 72(8), 923-932.

Chen, Q., Gong, P., Baldocchi, D., & Tian, Y. Q. (2007). Estimating basal area and stem volume for individual trees from lidar data. Photogrammetric Engineering and Remote Sensing, 73(12), 1355-1365.

Chirici, G., McRoberts, R. E., Winter, S., Bertini, R., Brändli, U.-B., Asensio, I. A., Marchetti, M. (2012). National forest inventory contributions to forest biodiversity monitoring. Forest Science, 58(3), 257-268.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Heimann, M. (2014). Carbon and other biogeochemical cycles Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (pp. 465-570): Cambridge University Press.

Clark, M. L., Roberts, D. A., Ewel, J. J., & Clark, D. B. (2011). Estimation of tropical rain forest aboveground biomass with small-footprint lidar and hyperspectral sensors. Remote Sensing of Environment, 115(11), 2931-2942.

Cole, T. G., & Ewel, J. J. (2006). Allometric equations for four valuable tropical tree species. Forest Ecology and Management, 229(1-3), 351-360. doi: 10.1016/j.foreco.2006.04.017

Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. IEEE Transactions on pattern analysis and machine intelligence, 24(5), 603-619.

Condit, R. (1995). Research in Large, Long-Term Tropical Forest Plots. Trends in Ecology & Evolution, 10(1), 18-22. doi:Doi 10.1016/S0169-5347(00)88955-7

Cook, R. D., & Weisberg, S. (1982). Residuals and influence in regression. New York: Chapman and Hall.

Corona, P., Fattorini, L., Franceschi, S., Chirici, G., Maselli, F., & Secondi, L. (2014). Mapping by spatial predictors exploiting remotely sensed and ground data: A comparative design-based perspective. Remote Sensing of Environment, 152, 29-37.

Crookston, N. L., & Finley, A. O. (2008). yaImpute: An R package for kNN imputation.

Curtin, R. A. (1964). Stand Density and the Relationship of Crown Width to Diameter and Height in Eucalyptus Obliqua. Australian Forestry, 28(2), 91-105. doi:10.1080/00049158.1964.10675358.

Dalponte, M., Orka, H. O., Ene, L. T., Gobakken, T., & Naesset, E. (2014). Tree crown delineation and tree species classification in boreal forests using hyperspectral and ALS data. Remote Sensing of Environment, 140, 306-317. doi:10.1016/j.rse.2013.09.006

Davies, S. J., Noor, N. S. M., LaFrankie, J. V., & Ashton, P. S. (2003). The trees of Pasoh Forest: stand structure and floristic composition of the 50-ha forest research plot Pasoh (pp. 35-50): Springer.

Dawkins, H. C. (1963). Crown diameters: Their relation to bole diameter in tropical forest trees. The Commonwealth Forestry Review, 42(4 (114)), 318-333.

Dean, T. J., Cao, Q. V., Roberts, S. D., & Evans, D. L. (2009). Measuring heights to crown base and crown median with LiDAR in a mature, even-aged loblolly pine stand. Forest Ecology and Management, 257(1), 126-133. doi:10.1016/j.foreco.2008.08.024

Dixon, R. K., Solomon, A. M., Brown, S., Houghton, R. A., Trexier, M. C., & Wisniewski, J. (1994). Carbon pools and flux of global forest ecosystems. Science, 263(5144), 185-190. doi:10.1126/science.263.5144.18510.1016/S0034-4257(02)00013-5

Dong, P. (2009). Characterization of individual tree crowns using three-dimensional shape signatures derived from LiDAR data. International journal of remote sensing, 30(24), 6621-6628.

Drake, J. B., Dubayah, R. O., Knox, R. G., Clark, D. B., & Blair, J. B. (2002). Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest. Remote Sensing of Environment, 81(2-3), 378-392. doi:Pii S0034-4257(02)00013-5

Du, L., Zhou, T., Zou, Z. H., Zhao, X., Huang, K. C., & Wu, H. (2014). Mapping Forest Biomass Using Remote Sensing and National Forest Inventory in China. Forests, 5(6), 1267-1283. doi:10.3390/f5061267

Du, S., Wang, Q., & Guo, L. (2014). Spatially varying relationships between land-cover change and driving factors at multiple sampling scales. J Environ Manage, 137, 101-110. doi:10.1016/j.jenvman.2014.01.037

Dubayah, R. O., & Drake, J. B. (2000). Lidar remote sensing for forestry. Journal of Forestry, 98(6), 44-46.

Duncanson, L. I., Cook, B. D., Hurtt, G. C., & Dubayah, R. O. (2014). An efficient, multi-layered crown delineation algorithm for mapping individual tree structure across multiple ecosystems. Remote Sensing of Environment, 154, 378-386. doi:10.1016/j.rse.2013.07.044

Ene, L. T., Naesset, E., Gobakken, T., Gregoire, T. G., Stahl, G., & Nelson, R. (2012). Assessing the accuracy of regional LiDAR-based biomass estimation using a

simulation approach. Remote Sensing of Environment, 123, 579-592. doi:10.1016/j.rse.2012.04.017

Enquist, B. J., Allen, A. P., Brown, J. H., Gillooly, J. F., Kerkhoff, A. J., Niklas, K. J., West, G. B. (2007). Biological scaling: does the exception prove the rule? Nature, 445(7127), E9-E10.

Eskelson, B. N., Temesgen, H., Lemay, V., Barrett, T. M., Crookston, N. L., & Hudak, A. T. (2009). The roles of nearest neighbor methods in imputing missing data in forest inventory and monitoring databases. Scandinavian Journal of Forest Research, 24(3), 235-246.

Evans, J. S., Hudak, A. T., Faux, R., & Smith, A. (2009). Discrete return lidar in natural resources: Recommendations for project planning, data processing, and deliverables. Remote Sensing, 1(4), 776-794.

Falkowski, M. J., Hudak, A. T., Crookston, N. L., Gessler, P. E., Uebler, E. H., & Smith, A. M. S. (2010). Landscape-scale parameterization of a tree-level forest growth model: a k-nearest neighbor imputation approach incorporating LiDAR data. Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere, 40(2), 184-199. doi:10.1139/X09-183.

Falkowski, M. J., Smith, A. M. S., Hudak, A. T., Gessler, P. E., Vierling, L. A., & Crookston, N. L. (2006). Automated estimation of individual conifer tree height and crown diameter via two-dimensional spatial wavelet analysis of lidar data. Canadian Journal of Remote Sensing, 32(2), 153-161.

FAO. (2010). Global Forest Resources Assessment 2010. FAO, Rome, Italy.

FAO. (2011). State of the world's forest 2011. FAO, Rome, Italy.

Faraway, J. J. (2014). Linear models with R: CRC press.

Fassnacht, F. E., Hartig, F., Latifi, H., Berger, C., Hernandez, J., Corvalan, P., & Koch, B. (2014). Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass. Remote Sensing of Environment, 154, 102-114. doi:10.1016/j.rse.2014.07.028

Fayad, I., Baghdadi, N., Bailly, J. S., Barbier, N., Gond, V., El Hajj, M., Bourgine, B. (2014). Canopy Height Estimation in French Guiana with LiDAR ICESat/GLAS Data Using Principal Component Analysis and Random Forest Regressions. Remote Sensing, 6(12), 11883-11914. doi:10.3390/rs61211883

Feldpausch, T. R., Banin, L., Phillips, O. L., Baker, T. R., Lewis, S. L., Quesada, C. A.,Lloyd, J. (2011). Height-diameter allometry of tropical forest trees. Biogeosciences, 8(5), 1081-1106. doi:10.5194/bg-8-1081-2011

Ferraz, A., Bretar, F., Jacquemoud, S., Gonçalves, G., & Pereira, L. (2010). 3D segmentation of forest structure using a mean-shift based algorithm. Paper presented at the Image Processing (ICIP), 2010 17th IEEE International Conference on.

Ferraz, A., Bretar, F., Jacquemoud, S., Goncalves, G., Pereira, L., Tome, M., & Soares, P. (2012). 3-D mapping of a multi-layered Mediterranean forest using ALS data. Remote Sensing of Environment, 121, 210-223. doi:10.1016/j.rse.2012.01.020

Ferraz, A., Saatchi, S., Mallet, C., & Meyer, V. (2016). Lidar detection of individual tree size in tropical forests. Remote Sensing of Environment, 183, 318-333. doi:10.1016/j.rse.2016.05.028

Fletcher, C., Abrams, M., Ibrahim, S., Musa, S., & Kassim, A. R. (2012). Fresh Perspecfives on Malaysia's Pasoh Forest Reserve and Climate Change.

Flewelling, J. W. (2009). Forest inventory predictions from individual tree crowns: regression modeling within a sample framework. Paper presented at the Proceedings of the Eighth Annual Forest Inventory and Analysis Symposium 2006, Monterey, CA. Gen.

Fox, J. C., Ades, P. K., & Bi, H. (2001). Stochastic structure and individual-tree growth models. Forest Ecology and Management, 154(1), 261-276.

Friedlingstein, P., & Prentice, I. C. (2010). Carbon–climate feedbacks: a review of model and observation based estimates. Current Opinion in Environmental Sustainability, 2(4), 251-257.

FRIM, Forest Research Institute Malaysia. (2011). Research on Development of Carbon Monitoring Methodology for REDD+ in Malaysia. Annual Report, March 30, 2011.

FRIM, Forest Research Institute Malaysia. (2012). Research on Development of Carbon Monitoring Methodology for REDD+ in Malaysia. Annual Report, 2012.

Frim.gov.my. (2018). Forest Research Institute Malaysia » FRIM-FDPM collaborate in R&D on remote sensing for REDD+ and Climate. [online] Available at: https://www.frim.gov.my/en/frim-fdpm-collaborate-in-rd-on-remote-sensing-for-redd-and-climate/ [Accessed 18 Mar. 2018]

Gagliasso, D., Hummel, S., & Temesgen, H. (2014). A comparison of selected parametric and non-parametric imputation methods for estimating forest biomass and basal area. Open Journal of Forestry, 4(1), 42.

García, M., Riaño, D., Chuvieco, E., & Danson, F. M. (2010). Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data. Remote Sensing of Environment, 114(4), 816-830.

Garcia-Gutierrez, J., Gonzalez-Ferreiro, E., Riquelme-Santos, J. C., Miranda, D., Dieguez-Aranda, U., & Navarro-Cerrillo, R. M. (2014). Evolutionary feature selection to estimate forest stand variables using LiDAR. International Journal of Applied Earth Observation and Geoinformation, 26, 119-131. doi:10.1016/j.jag.2013.06.005

Gatziolis, D., & Andersen, H.-E. (2008). A guide to LIDAR data acquisition and processing for the forests of the Pacific Northwest.

Gering, L. R., & May, D. M. (1995). The relationship of diameter at breast height and crown diameter for four species groups in Hardin County, Tennessee. Southern Journal of Applied Forestry, 19(4), 177-181.

Gibbs, H. K., Brown, S., Niles, J. O., & Foley, J. A. (2007). Monitoring and estimating tropical forest carbon stocks: making REDD a reality. Environmental Research Letters, 2(4). doi:Artn 045023

Gibbs, H. K., Ruesch, A. S., Achard, F., Clayton, M. K., Holmgren, P., Ramankutty, N., & Foley, J. A. (2010). Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. Proceedings of the National Academy of Sciences, 107(38), 16732-16737.

Gillis, M. D., & Leckie, D. G. (1993). Forest inventory mapping procedures across Canada (Vol. 114).

Gleason, C. J., & Im, J. (2012). Forest biomass estimation from airborne LiDAR data using machine learning approaches. Remote Sensing of Environment, 125, 80-91.

Gobakken, T., Bollandsas, O. M., & Naesset, E. (2015). Comparing biophysical forest characteristics estimated from photogrammetric matching of aerial images and airborne laser scanning data. Scandinavian Journal of Forest Research, 30(1), 73-86. doi:10.1080/02827581.2014.961954

Gobakken, T., Naesset, E., Nelson, R., Bollandsas, O. M., Gregoire, T. G., Stahl, G., Astrup, R. (2012). Estimating biomass in Hedmark County, Norway using national forest inventory field plots and airborne laser scanning. Remote Sensing of Environment, 123, 443-456. doi:10.1016/j.rse.2012.01.025.

Goerndt, M. E., Monleon, V. J., & Temesgen, H. (2010). Relating Forest Attributes with Area- and Tree-Based Light Detection and Ranging Metrics for Western Oregon. Western Journal of Applied Forestry, 25(3), 105-111.

Goetz, S., & Dubayah, R. (2011). Advances in remote sensing technology and implications for measuring and monitoring forest carbon stocks and change. Carbon Management, 2(3), 231-244. doi:10.4155/Cmt.11.18

Goodman, R. C., Phillips, O. L., & Baker, T. R. (2014). The importance of crown dimensions to improve tropical tree biomass estimates. Ecological Applications, 24(4), 680-698. doi:10.1890/13-0070.1

Gorte, B. (2002). Segmentation of TIN-structured surface models. International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences, 34(4), 465-469.

Goutte, C., & Gaussier, E. (2005). A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In *ECIR* (Vol. 5, pp. 345-359).

Grace, J., & Meir, P. (2009). Tropical Rain Forests as Old-Growth Forests. Old-Growth Forests: Function, Fate and Value, 207, 391-408. doi:10.1007/978-3-540-92706-8_17

Grace, J., Mitchard, E., & Gloor, E. (2014). Perturbations in the carbon budget of the tropics. Global Change Biology, 20(10), 3238-3255.

Gregoire, T. G., & Valentine, H. T. (2008). Sampling Strategies for Natural and Environmental Resources: Chapman & Hall/CRC.

Gupta, S., Weinacker, H., & Koch, B. (2010). Comparative Analysis of Clustering-Based Approaches for 3-D Single Tree Detection Using Airborne Fullwave Lidar Data. Remote Sensing, 2(4), 968-989. doi:10.3390/rs2040968

Hadji, I., & Nabelek, D. (n.d.). Clustering algorithms used in 3D scene segmentation.

Hall, S. A., Burke, I. C., Box, D. O., Kaufmann, M. R., & Stoker, J. M. (2005). Estimating stand structure using discrete-return lidar: an example from low density, fire prone ponderosa pine forests. Forest Ecology and Management, 208(1-3), 189-209. doi:10.1016/j.foreco.2004.12.001

Hamdan, O., Hasmadi, I. M., Aziz, H. K., Norizah, K., & Zulhaidi, M. H. (2015). L-band saturation level for aboveground biomass of dipterocarp forests in peninsular Malaysia. Journal of Tropical Forest Science, 388-399.

Hamdan, O., Mohd Azahari, F., Abd Rahman, K., Nor Azura, O., & Nurul Dasani, A. D. (2015). Airborne LiDAR for estimating aboveground biomass in dipterocarp forests of Malaysia. Paper presented at the ACRS 2015 - 36th Asian Conference on Remote Sensing, Quezon City, Manila.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Townshend, J. R. G. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. Science, 342(6160), 850-853. doi:10.1126/science.1244693.

Harrell Jr, F. E. (2015). Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis: Springer.

Hauglin, M., Dibdiakova, J., Gobakken, T., & Naesset, E. (2013). Estimating single-tree branch biomass of Norway spruce by airborne laser scanning. ISPRS Journal of Photogrammetry and Remote Sensing, 79, 147-156. doi:10.1016/j.isprsjprs.2013.02.013

Hauglin, M., Lien, V., Naesset, E., & Gobakken, T. (2014). Geo-referencing forest field plots by co-registration of terrestrial and airborne laser scanning data. International Journal of Remote Sensing, 35(9), 3135-3149. doi:10.1080/01431161.2014.903440

He, Q. S., Chen, E. X., An, R., & Li, Y. (2013). Above-Ground Biomass and Biomass Components Estimation Using LiDAR Data in a Coniferous Forest. Forests, 4(4), 984-1002. doi:10.3390/f4040984

Heinzel, J. N., Weinacker, H., & Koch, B. (2011). Prior-knowledge-based single-tree extraction. International Journal of Remote Sensing, 32(17), 4999-5020. doi:10.1080/01431161.2010.494633

Henry, M., Besnard, A., Asante, W. A., Eshun, J., Adu-Bredu, S., Valentini, R., . . . Saint-Andre, L. (2010). Wood density, phytomass variations within and among trees, and allometric equations in a tropical rainforest of Africa. Forest Ecology and Management, 260(8), 1375-1388. doi:10.1016/j.foreco.2010.07.040

Heritage, C. o. (2015). FRIM Tropical Man-Made Forest Research Park.

Hollaus, M., Dorigo, W., Wagner, W., Schadauer, K., Höfle, B., & Maier, B. (2009). Operational wide-area stem volume estimation based on airborne laser

scanning and national forest inventory data. International Journal of Remote Sensing, 30(19), 5159-5175.

Holmgren, J., Nilsson, M., & Olsson, H. (2003). Simulating the effects of lidar scanning angle for estimation of mean tree height and canopy closure. Canadian Journal of Remote Sensing, 29(5), 623-632.

Holmgren, J., Barth, A., Larsson, H., & Olsson, H. (2012). Prediction of stem attributes by combining airborne laser scanning and measurements from harvesters. Silva Fenn, 46(2), 227-239.

Houghton, R. A. (2005). Tropical deforestation as a source of greenhouse gas emissions. In the Woods Hole research centre In: P. Moutinho and S. Schwarzman (eds) (Ed.), Tropical Deforestation and Climate Change (pp. 13). Washington: Instituto de Pesquisa Ambiental da Amazônia Amazon Institute for Environmental Research.

Houghton, R. A., House, J., Pongratz, J., Van Der Werf, G., DeFries, R., Hansen, M., Ramankutty, N. (2012). Carbon emissions from land use and land-cover change. Biogeosciences, 9(12), 5125-5142.

Houghton, R. A., Lawrence, K. T., Hackler, J. L., & Brown, S. (2001). The spatial distribution of forest biomass in the Brazilian Amazon: a comparison of estimates. Global Change Biology, 7(7), 731-746.

Hu, B., Li, J., Jing, L., & Judah, A. (2014). Improving the efficiency and accuracy of individual tree crown delineation from high-density LiDAR data. International Journal of Applied Earth Observation and Geoinformation, 26, 145-155.

Hu, X. B., Chen, W., & Xu, W. Y. (2017). Adaptive Mean Shift-Based Identification of Individual Trees Using Airborne LiDAR Data. Remote Sensing, 9(2). doi:ARTN 148

Hu, X., & Xie, Y. (2016). Segmentation and clustering of 3D forest point cloud using mean shift algorithms. Paper presented at the Proceedings of the 4th International Conference on Machinery, Materials and Computing Technology (ICMMCT 2016), Hangzhou, China.

Hudak, A. T., Crookston, N. L., Evans, J. S., Falkowski, M. J., Smith, A. M., Gessler, P. E., & Morgan, P. (2006). Regression modeling and mapping of coniferous forest basal area and tree density from discrete-return lidar and multispectral satellite data. Canadian Journal of Remote Sensing, 32(2), 126-138.

Hudak, A. T., Crookston, N. L., Evans, J. S., Hall, D. E., & Falkowski, M. J. (2008). Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. Remote Sensing of Environment, 112(5), 2232-2245. doi:10.1016/j.rse.2007.10.009

Hudak, A. T., Crookston, N. L., Evans, J. S., Hall, D. E., & Falkowski, M. J. (2008). Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. Remote Sensing of Environment, 112(5), 2232-2245. doi:10.1016/j.rse.2007.10.009

Hudak, A. T., Crookston, N. L., Evans, J. S., Hall, D. E., & Falkowski, M. J. (2008). Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. Remote Sensing of Environment, 112(5), 2232-2245. doi:10.1016/j.rse.2007.10.009

Hudak, A. T., Haren, A. T., Crookston, N. L., Liebermann, R. J., & Ohmann, J. L. (2014). Imputing Forest Structure Attributes from Stand Inventory and Remotely Sensed Data in Western Oregon, USA. Forest Science, 60(2), 253-269. doi:10.5849/forsci.12-101

Hudak, A. T., Haren, A. T., Crookston, N. L., Liebermann, R. J., & Ohmann, J. L. (2014). Imputing Forest Structure Attributes from Stand Inventory and Remotely Sensed Data in Western Oregon, USA. Forest Science, 60(2), 253-269. doi:10.5849/forsci.12-101

Hudak, A. T., Strand, E. K., Vierling, L. A., Byrne, J. C., Eitel, J. U. H., Martinuzzi, S., & Falkowski, M. J. (2012). Quantifying aboveground forest carbon pools and fluxes from repeat LiDAR surveys. Remote Sensing of Environment, 123, 25-40. doi:10.1016/j.rse.2012.02.023.

Husch, B., Beers, T. W., & Kershaw Jr, J. A. (2002). Forest mensuration: John Wiley & Sons.

Hyde, P., Nelson, R., Kimes, D., & Levine, E. (2007). Exploring LiDAR–RaDAR synergy—predicting aboveground biomass in a southwestern ponderosa pine forest using LiDAR, SAR and InSAR. Remote Sensing of Environment, 106(1), 28-38.

Hyyppä, J., & Inkinen, M. (1999). Detecting and estimating attributes for single trees using laser scanner. The Photogrammetric Journal of Finland, 16(2), 27-42. doi:citeulike-article-id:12239821

Hyyppa, J., Hyyppa, H., Leckie, D., Gougeon, F., Yu, X., & Maltamo, M. (2008). Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. International Journal of Remote Sensing, 29(5), 1339-1366. doi:10.1080/01431160701736489.

Hyyppa, J., Kelle, O., Lehikoinen, M., & Inkinen, M. (2001). A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. Ieee Transactions on Geoscience and Remote Sensing, 39(5), 969-975. doi:Doi 10.1109/36.921414

Ioki, K., Tsuyuki, S., Hirata, Y., Phua, M. H., Wong, W. V. C., Ling, Z. Y., . . . Takao, G. (2014). Estimating above-ground biomass of tropical rainforest of different degradation levels in Northern Borneo using airborne LiDAR. Forest Ecology and Management, 328, 335-341. doi:10.1016/j.foreco.2014.06.003

IPCC. (2003). Good Practice Guidance for Land Use, Land-Use Change and Forestry. Institute for Global Environmental Strategies (IGES), Hayama, Japan, ISBN 4-88788-003-0.

Ismail, M. H., & Manaf, M. S. (2011). The potential of LiDAR application in Malaysia. International Journal of Remote Sensing Applications, 1(1), 1-5.

International Organization for Standardization. ISO. (2012). Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions. Retrieved from https://www.iso.org/obp/ui/#iso:std:iso:5725:1:ed1:v1:en10.1088/17489326/2/4/04502310.1088/1748-9326/4/3/03400910.1175/2009ei278.110.3390/rs804033310.3390/rs9020148

Iverson, L. R., Cook, E. A., & Graham, R. L. (1989). A Technique for Extrapolating and Validating Forest Cover across Large Regions - Calibrating Avhrr Data with Tm Data. International Journal of Remote Sensing, 10(11), 1805-1812.

Jakubowski, M. K., Li, W., Guo, Q., & Kelly, M. (2013). Delineating individual trees from LiDAR data: A comparison of vector-and raster-based segmentation approaches. Remote Sensing, 5(9), 4163-4186.

Jakubowski, M. K., Li, W., Guo, Q., & Kelly, M. (2013). Delineating individual trees from LiDAR data: A comparison of vector-and raster-based segmentation approaches. Remote Sensing, 5(9), 4163-4186.

Jaskierniak, D., Lane, P. N., Robinson, A., & Lucieer, A. (2011). Extracting LiDAR indices to characterise multilayered forest structure using mixture distribution functions. Remote Sensing of Environment, 115(2), 573-585.

Jing, L. H., Hu, B. X., Noland, T., & Li, J. L. (2012). An individual tree crown delineation method based on multi-scale segmentation of imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 70, 88-98. doi:10.1016/j.isprsjprs.2012.04.003

Johnson, K. D., Birdsey, R., Finley, A. O., Swantaran, A., Dubayah, R., Wayson, C., & Riemann, R. (2014). Integrating forest inventory and analysis data into a LIDAR-based carbon monitoring system. Carbon Balance and Management, 9(1), 1-11. doi:10.1186/1750-0680-9-3

Jubanski, J., Ballhorn, U., Kronseder, K., Franke, J., & Siegert, F. (2013). Detection of large above-ground biomass variability in lowland forest ecosystems by airborne LiDAR. Biogeosciences, 10(6), 3917-3930. doi:10.5194/bg-10-3917-2013

Kaartinen, H., Hyyppa, J., Yu, X. W., Vastaranta, M., Hyyppa, H., Kukko, A., Wu, J. C. (2012). An International Comparison of Individual Tree Detection and Extraction Using Airborne Laser Scanning. Remote Sensing, 4(4), 950-974. doi:10.3390/rs4040950

Kankare, V., Vastaranta, M., Holopainen, M., Raty, M., Yu, X. W., Hyyppa, J.,Viitala, R. (2013). Retrieval of Forest Aboveground Biomass and Stem Volume with Airborne Scanning LiDAR. Remote Sensing, 5(5), 2257-2274. doi:10.3390/rs5052257

Kassim, A. R. (2015). FRIM Tropical Man-Made Forest Research Park.

Kato, R., Tadaki, Y., & Ogawa, H. (1978). Plant biomass and growth increment studies in Pasoh Forest. Malayan Nature Journal.

Keith, H., Mackey, B. G., & Lindenmayer, D. B. (2009). Re-evaluation of forest biomass carbon stocks and lessons from the world's most carbon-dense forests. Proceedings of the National Academy of Sciences of the United States of America, 106(28), 11635-11640. doi:10.1073/pnas.0901970106

Keller, M., Palace, M., & Hurtt, G. (2001). Biomass estimation in the Tapajos National Forest, Brazil - Examination of sampling and allometric uncertainties. Forest Ecology and Management, 154(3), 371-382. doi:Doi 10.1016/S0378-1127(01)00509-6

Kenzo, T., Ichie, T., Hattori, D., Itioka, T., Handa, C., Ohkubo, T.,Ninomiya, I. (2009). Development of allometric relationships for accurate estimation of above- and below-ground biomass in tropical secondary forests in Sarawak, Malaysia. Journal of Tropical Ecology, 25, 371-386.

Ketterings, Q. M., Coe, R., van Noordwijk, M., Ambagau, Y., & Palm, C. A. (2001). Reducing uncertainty in the use of allometric biomass equations for predicting above-ground tree biomass in mixed secondary forests. Forest Ecology and Management, 146(1-3), 199-209.

Khosravipour, A., Skidmore, A. K., Isenburg, M., Wang, T. J., & Hussin, Y. A. (2014). Generating Pit-free Canopy Height Models from Airborne Lidar. Photogrammetric Engineering and Remote Sensing, 80(9), 863-872.

Kim, Y., Yang, Z. Q., Cohen, W. B., Pflugmacher, D., Lauver, C. L., & Vankat, J. L. (2009). Distinguishing between live and dead standing tree biomass on the North Rim of Grand Canyon National Park, USA using small-footprint lidar data. Remote Sensing of Environment, 113(11), 2499-2510. doi:10.1016/j.rse.2009.07.010

Koch, B., Heyder, U., & Weinacker, H. (2006). Detection of individual tree crowns in airborne lidar data. Photogrammetric Engineering and Remote Sensing, 72(4), 357-363.

Koch, B., Kattenborn, T., Straub, C., & Vauhkonen, J. (2014). Segmentation of Forest to Tree Objects. Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies, 27, 89-112. doi:10.1007/978-94-017-8663-8_5

Koukoulas, S., & Blackburn, G. A. (2005). Mapping individual tree location, height and species in broadleaved deciduous forest using airborne LIDAR and multi-spectral remotely sensed data. International Journal of Remote Sensing, 26(3), 431-455. doi:10.1080/0143116042000298289

Kramer, H. A., Collins, B. M., Gallagher, C. V., J Keane, J., Stephens, S. L., & Kelly, M. (2016). Accessible light detection and ranging: estimating large tree density for habitat identification. Ecosphere, 7(12).

Kraus, K., & Mikhail, E. (1972). Linear least-squares interpolation. Amer Soc Photogrammetry 5410 Grosvenor Lane Suite 210, Bethesda, MD 20814-2160.

Kraus, K., & Pfeifer, N. (1998). Determination of terrain models in wooded areas with airborne laser scanner data. ISPRS Journal of Photogrammetry and Remote Sensing, 53(4), 193-203.

Kraus, K., & Pfeifer, N. (2001). Advanced DTM generation from LIDAR data. International Archives Of Photogrammetry Remote Sensing And Spatial Information Sciences, 34(3/W4), 23-30.

Kwak, D. A., Lee, W. K., Lee, J. H., Biging, G. S., & Gong, P. (2007). Detection of individual trees and estimation of tree height using LiDAR data. Journal of Forest Research, 12(6), 425-434. doi:10.1007/s10310-007-0041-9

Kwan, W. Y. (1966). Crown Diameter/bole diameter relationship of Dyera costulata and its planting distance and crop density. Malayan Forester, 29(3), 163-169.

Lamas, T., & Eriksson, L. (2003). Analysis and planning systems for multiresource, sustainable forestry: the Heureka research programme at SLU. Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere, 33(3), 500-508. doi:10.1139/X02-213

Langner, A., Miettinen, J., & Siegert, F. (2007). Land cover change 2002-2005 in Borneo and the role of fire derived from MODIS imagery. Global Change Biology, 13(11), 2329-2340.

Latifi, H., & Koch, B. (2012). Evaluation of most similar neighbour and random forest methods for imputing forest inventory variables using data from target and auxiliary stands. International Journal of Remote Sensing, 33(21), 6668-6694.

Latifi, H., Nothdurft, A., & Koch, B. (2010). Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: application of multiple optical/LiDAR-derived predictors. Forestry, 83(4), 395-407.

Laurance, W. F. (2007). Forest destruction in tropical Asia. Current Science, 1544-1550.

Läuter, H. (1985). Cook, R. D., S. Weisberg: Residuals and influence in regression. Chapman and Hall, New York — London 1982. VIII, 229 pp., £ 12. Biometrical Journal, 27(1), 80-80. doi:10.1002/bimj.4710270110

Le Toan, T., Quegan, S., Davidson, M. W. J., Balzter, H., Paillou, P., Papathanassiou, K., . . . Ulander, L. (2011). The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. Remote Sensing of Environment, 115(11), 2850-2860. doi:10.1016/j.rse.2011.03.020

Leckie, D., Gougeon, F., Hill, D., Quinn, R., Armstrong, L., & Shreenan, R. (2003). Combined high-density lidar and multispectral imagery for individual tree crown analysis. Canadian Journal of Remote Sensing, 29(5), 633-649.

Lee, H., Slatton, K. C., Roth, B. E., & Cropper, W. P. (2010). Adaptive clustering of airborne LiDAR data to segment individual tree crowns in managed pine forests. International Journal of Remote Sensing, 31(1), 117-139. doi:10.1080/01431160902882561

Lee, J., Coomes, D., Schonlieb, C.-B., Cai, X., Lellmann, J., Dalponte, M., . . . Morecroft, M. (2017). A graph cut approach to 3D tree delineation, using integrated airborne LiDAR and hyperspectral imagery. arXiv preprint arXiv:1701.06715.

Lefsky, M. A., Cohen, W. B., Acker, S. A., Parker, G. G., Spies, T. A., & Harding, D. (1999). Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. Remote Sensing of Environment, 70(3), 339-361. doi:Doi 10.1016/S0034-4257(99)00052-8

Lefsky, M. A., Cohen, W. B., Harding, D. J., Parker, G. G., Acker, S. A., & Gower, S. T. (2002). Lidar remote sensing of above-ground biomass in three biomes. Global Ecology and Biogeography, 11(5), 393-399. doi:DOI 10.1046/j.1466-822x.2002.00303.x

Lefsky, M. A., Hudak, A. T., Cohen, W. B., & Acker, S. A. (2005). Patterns of covariance between forest stand and canopy structure in the Pacific Northwest. Remote Sensing of Environment, 95(4), 517-531.

Leitold, V., Keller, M., Morton, D. C., Cook, B. D., & Shimabukuro, Y. E. (2015). Airborne lidar-based estimates of tropical forest structure in complex terrain: opportunities and trade-offs for REDD+. Carbon Balance and Management, 10(1), 3.

LeMay, V., Maedel, J., & Coops, N. C. (2008). Estimating stand structural details using nearest neighbor analyses to link ground data, forest cover maps, and Landsat imagery. Remote Sensing of Environment, 112(5), 2578-2591.

Li, H., Mausel, P., Brondizio, E., & Deardorff, D. (2010). A framework for creating and validating a non-linear spectrum-biomass model to estimate the secondary succession biomass in moist tropical forests. ISPRS Journal of Photogrammetry and Remote Sensing, 65(2), 241-254. doi:10.1016/j.isprsjprs.2010.01.002

Li, J., Jiang, Y. F., Yang, C. W., Huang, Q. Y., & Rice, M. (2013). Visualizing 3D/4D environmental data using many-core graphics processing units (GPUs) and multi-core central processing units (CPUs). Computers & Geosciences, 59, 78-89. doi:10.1016/j.cageo.2013.04.029

Li, M., Im, J., Quackenbush, L. J., & Liu, T. (2014). Forest Biomass and Carbon Stock Quantification Using Airborne LiDAR Data: A Case Study Over Huntington Wildlife Forest in the Adirondack Park. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7(7), 3143-3156. doi:10.1109/Jstars.2014.2304642.

Li, W. K., Guo, Q. H., Jakubowski, M. K., & Kelly, M. (2012). A New Method for Segmenting Individual Trees from the Lidar Point Cloud. Photogrammetric Engineering and Remote Sensing, 78(1), 75-84.

Li, Y. Z., Andersen, H. E., & McGaughey, R. (2008). A Comparison of Statistical Methods for Estimating Forest Biomass from Light Detection and Ranging Data. Western Journal of Applied Forestry, 23(4), 223-231.

Lichstein, J. W., Dushoff, J., Ogle, K., Chen, A., Purves, D. W., Caspersen, J. P., & Pacala, S. W. (2010). Unlocking the forest inventory data: relating individual tree performance to unmeasured environmental factors. *Ecological Applications*, *20*(3), 684-699.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. R news, 2(3), 18-22.

Lim, K. S., & Treitz, P. M. (2004). Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. Scandinavian Journal of Forest Research, 19(6), 558-570. doi:10.1080/02827580410019490

Lim, K., Treitz, P., Wulder, M., St-Onge, B., & Flood, M. (2003). LiDAR remote sensing of forest structure. Progress in Physical Geography, 27(1), 88-106. doi:10.1191/0309133303pp360ra.

Lindberg, E., & Holmgren, J. (2017). Individual Tree Crown Methods for 3D Data from Remote Sensing. Current Forestry Reports, 3(1), 19-31.

Lindberg, E., Holmgren, J., Olofsson, K., Olsson, H., & Wallerman, J. (2008). Estimation of tree lists from airborne laser scanning data using a combination of analysis on single tree and raster cell level. Paper presented at the Proceedings of the SilviLaser 2008 Conference, Edinburgh, UK. Available from geography. swan. ac. uk/silvilaser/papers/poster_papers/Lindberg. pdf [accessed 1 June 2010].

Liu, Q., Jing, L., Li, Y., Tang, Y., Li, H., & Lin, Q. (2016). A tree canopy height delineation method based on Morphological Reconstruction—Open Crown Decomposition. Paper presented at the IOP Conference Series: Earth and Environmental Science.

Lloyd, S. (1982). Least squares quantization in PCM. IEEE transactions on information theory, 28(2), 129-137.

Loetsch, F., Zöhrer, F., & Haller, K. (1973). Forest inventory. Vol. II: BLV Verlagsgesellschaft, München. ISBN.

Lu, D. S. (2006). The potential and challenge of remote sensing-based biomass estimation. International Journal of Remote Sensing, 27(7), 1297-1328.

Lu, D., Chen, Q., Wang, G., Moran, E., Batistella, M., Zhang, M., . . . Saah, D. (2012). Aboveground Forest Biomass Estimation with Landsat and LiDAR Data and Uncertainty Analysis of the Estimates. International Journal of Forestry Research, 2012, 1-16. doi:10.1155/2012/436537

Lucas, R. M., Cronin, N., Lee, A., Moghaddam, M., Witte, C., & Tickle, P. (2006). Empirical relationships between AIRSAR backscatter and LiDAR-derived forest biomass, queensland, Australia. Remote Sensing of Environment, 100(3), 407-425. doi:10.1016/j.rse.2005.10.1019

Magnussen, S., & Boudewyn, P. (1998). Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. Canadian Journal of Forest Research, 28(7), 1016-1031.

Magnussen, S., Eggermont, P., & LaRiccia, V. N. (1999). Recovering tree heights from airborne laser scanner data. Forest Science, 45(3), 407-422.

Magnusson, M. (2006). Evaluation of remote sensing techniques for estimation of forest variables at stand level (Vol. 2006).

Mallet, C., & Bretar, F. (2009). Full-waveform topographic lidar: State-of-the-art. ISPRS Journal of Photogrammetry and Remote Sensing, 64(1), 1-16. doi:10.1016/j.isprsjprs.2008.09.007

Maltamo, M., Eerikainen, K., Packalen, P., & Hyyppa, J. (2006). Estimation of stem volume using laser scanning-based canopy height metrics. Forestry, 79(2), 217-229. doi:10.1093/forestry/cpl007

Maltamo, M., Mustonen, K., Hyyppa, J., Pitkanen, J., & Yu, X. (2004). The accuracy of estimating individual tree variables with airborne laser scanning in a boreal nature reserve. Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere, 34(9), 1791-1801. doi:10.1139/X04-055

Maltamo, M., Naesset, E., & Vauhkonen, J. (2014). Forestry Applications of Airborne Laser Scanning Concepts and Case Studies Preface. Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies, 27, V-V. doi:Book_Doi 10.1007/978-94-017-8663-8

Maltamo, M., Peuhkurinen, J., Malinen, J., Vauhkonen, J., Packalén, P., & Tokola, T. (2009). Predicting Tree Attributes and Quality Characteristics of Scots Pine Using Airborne Laser Scanning Data. Silva Fennica, 43(3), 507-521.

Manokaran, N. (1998). Effect, 34 years later, of selective logging in the lowland dipterocarp forest at Pasoh, Peninsular Malaysia, and implications on present day logging in the hill forests. Conservation, Management and Development of Forest Resources, Forest Research Institute, Kuala Lumpur, 41-60.

Manokaran, N., & LaFrankie Jr, J. (1990). Stand structure of Pasoh Forest Reserve, a lowland rain forest in Peninsular Malaysia. Journal of Tropical Forest Science, 14-24.

Mauya, E. W., Ene, L. T., Bollandsås, O. M., Gobakken, T., Næsset, E., Malimbwi, R. E., & Zahabu, E. (2015). Modelling aboveground forest biomass using airborne laser scanner data in the miombo woodlands of Tanzania. Carbon Balance and Management, 10(1), 28.

McGaughey, R. J. (2009). FUSION/LDV: Software for LIDAR data analysis and visualization. US Department of Agriculture, Forest Service, Pacific Northwest Research Station: Seattle, WA, USA, 123(2).

McGaughey, R. J. (2014). Fusion/LDV: Software for LiDAR data analysis and visualization version 3.50. Retrieved from http://forsys.cfr.washington.edu/fusion/FUSION_manual.pdf

McRoberts, R. E., Tomppo, E. O., & Naesset, E. (2010). Advances and emerging issues in national forest inventories. Scandinavian Journal of Forest Research, 25(4), 368-381. doi:10.1080/02827581.2010.496739

McRoberts, R. E., Tomppo, E. O., Finley, A. O., & Heikkinen, J. (2007). Estimating areal means and variances of forest attributes using the k-Nearest Neighbors technique and satellite imagery. Remote Sensing of Environment, 111(4), 466-480. doi:10.1016/j.rse.2007.04.002

Means, J. E., Acker, S. A., Fitt, B. J., Renslow, M., Emerson, L., & Hendrix, C. J. (2000). Predicting forest stand characteristics with airborne scanning lidar. Photogrammetric Engineering and Remote Sensing, 66(11), 1367-1372.

Means, J. E., Acker, S. A., Harding, D. J., Blair, J. B., Lefsky, M. A., Cohen, W. B., McKee, W. A. (1999). Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the Western Cascades of Oregon. Remote Sensing of Environment, 67(3), 298-308.

Merrick, M. J., Koprowski, J. L., & Wilcox, C. (2013). Into the third dimension: benefits of incorporating LiDAR data in wildlife habitat models.

Meyer, F., & Beucher, S. (1990). Morphological segmentation. Journal of visual communication and image representation, 1(1), 21-46.

Minh, D. H. T., Toan, T. L., Rocca, F., Tebaldini, S., d'Alessandro, M. M., & Villard, L. (2014). Relating P-Band Synthetic Aperture Radar Tomography to Tropical Forest Biomass. Ieee Transactions on Geoscience and Remote Sensing, 52(2), 967-979. doi:10.1109/Tgrs.2013.2246170

Mitchard, E. T. A., Saatchi, S. S., Gerard, F. F., Lewis, S. L., & Meir, P. (2009). Measuring Woody Encroachment along a Forest-Savanna Boundary in Central Africa. Earth Interactions, 13. doi:Artn 8

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to linear regression analysis (Vol. 821): John Wiley & Sons.

Morel, A. C., Saatchi, S. S., Malhi, Y., Berry, N. J., Banin, L., Burslem, D., . . . Ong, R. C. (2011). Estimating aboveground biomass in forest and oil palm plantation in Sabah, Malaysian Borneo using ALOS PALSAR data. Forest Ecology and Management, 262(9), 1786-1798.

Morsdorf, F., Meier, E., Allgöwer, B., & Nüesch, D. (2003). Clustering in airborne laser scanning raw data for segmentation of single trees. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 34(part 3), W13.

Morsdorf, F., Meier, E., Kötz, B., Itten, K. I., Dobbertin, M., & Allgöwer, B. (2004). LIDAR-based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management. Remote Sensing of Environment, 92(3), 353-362.

Morsdorf, F., Nichol, C., Malthus, T. J., Patenaude, G., & Woodhouse, I. H. (2008). Modelling multi-spectral LIDAR vegetation backscatter–assessing structural and physiological information content. SilviLaser 2008, 17-19.

Muinonen, E., Maltamo, M., Hyppanen, H., & Vainikainen, V. (2001). Forest stand characteristics estimation using a most similar neighbor approach and image spatial structure information. Remote Sensing of Environment, 78(3), 223-228. doi:Doi 10.1016/S0034-4257(01)00220-6

Naesset, E. (1997). Determination of mean tree height of forest stands using airborne laser scanner data. ISPRS Journal of Photogrammetry and Remote Sensing, 52(2), 49-56.

Naesset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sensing of Environment, 80(1), 88-99.

Naesset, E. (2014). Area-Based Inventory in Norway - From Innovation to an Operational Reality. Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies, 27, 215-240. doi:10.1007/978-94-017-8663-8_11

Naesset, E., & Gobakken, T. (2008). Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. Remote Sensing of Environment, 112(6), 3079-3090. doi:10.1016/j.rse.2008.03.004

Næsset, E., & Økland, T. (2002). Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. Remote Sensing of Environment, 79(1), 105-115.

Naesset, E., Bollandsas, O. M., & Gobakken, T. (2005). Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data. Remote Sensing of Environment, 94(4), 541-553.

Naesset, E., Bollandsas, O. M., Gobakken, T., Gregoire, T. G., & Stahl, G. (2013). Model-assisted estimation of change in forest biomass over an 11 year period in a sample survey supported by airborne LiDAR: A case study with post-stratification to provide "activity data". Remote Sensing of Environment, 128, 299-314. doi:10.1016/j.rse.2012.10.008

Nelson, R., Short, A., & Valenti, M. (2004). Measuring biomass and carbon in Delaware using an airborne profiling LIDAR. Scandinavian Journal of Forest Research, 19(6), 500-511. doi:10.1080/02827580410019508

Nelson, R., Valenti, M. A., Short, A., & Keller, C. (2003). A multiple resource inventory of Delaware using airborne laser data. AIBS Bulletin, 53(10), 981-992.

Niiyama, K., Kajimoto, T., Matsuura, Y., Yamashita, T., Matsuo, N., Yashiro, Y., Noor, N. S. (2010). Estimation of root biomass based on excavation of individual root systems in a primary dipterocarp forest in Pasoh Forest Reserve, Peninsular Malaysia. Journal of Tropical Ecology, 26, 271-284.

Niklas, K. J., & Enquist, B. J. (2001). Invariant scaling relationships for interspecific plant biomass production rates and body size. Proceedings of the National Academy of Sciences of the United States of America, 98(5), 2922-2927.

Nilsson, M. (1996). Estimation of tree heights and stand volume using an airborne lidar system. Remote Sensing of Environment, 56(1), 1-7.

Ohmann, J. L., & Gregory, M. J. (2002). Predictive mapping of forest composition and structure with direct gradient analysis and nearest-neighbor imputation in coastal Oregon, USA. Canadian Journal of Forest Research, 32(4), 725-741.

Okuda, T., Suzuki, M., Adachi, N., Yoshida, K., Niiyama, K., Noor, N. S. M., Hashim, M. (2003). Logging history and its impact on forest structure and species

composition in the Pasoh forest reserve—implications for the sustainable management of natural resources and landscapes Pasoh (pp. 15-34): Springer.

Omar, H. (2012, 2-4 April 2012). An Overview of REDD+ Activities in Malaysia. Paper presented at the 5th GEOSS-AP Symposium Miraikan, Tokyo, Japan.

Packalén, P., & Maltamo, M. (2007). The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. Remote Sensing of Environment, 109(3), 328-341.

Packalén, P., Temesgen, H., & Maltamo, M. (2012). Variable selection strategies for nearest neighbor imputation methods used in remote sensing based forest inventory. Canadian Journal of Remote Sensing, 38(5), 557-569.

Packalen, P., Vauhkonen, J., Kallio, E., Peuhkurinen, J., Pitkanen, J., Pippuri, I., . . . Maltamo, M. (2013). Predicting the spatial pattern of trees by airborne laser scanning. International Journal of Remote Sensing, 34(14), 5154-5165. doi:10.1080/01431161.2013.787501

Patenaude, G., Hill, R. A., Milne, R., Gaveau, D. L. A., Briggs, B. B. J., & Dawson, T. P. (2004). Quantifying forest above ground carbon content using LiDAR remote sensing. Remote Sensing of Environment, 93(3), 368-380. doi:10.1016/j.rse.2004.07.016.

Penner, M., Pitt, D., & Woods, M. (2013). Parametric vs. nonparametric LiDAR models for operational forest inventory in boreal Ontario. Canadian Journal of Remote Sensing, 39(5), 426-443.

Perez, J. W. (1970). Relation of crown diameter to stem diameter in forest of Puerto Rico, Dominic and Thailand A Tropical Rainforest – a study of irradiation and ecology at El Verde (pp. B105-B122). Odum: H.T; pigeon, R.F., eds: U.S. Atomic Energy Comission, Division of Technical In formation.

Persson, A., Holmgren, J., & Soderman, U. (2002). Detecting and measuring individual trees using an airborne laser scanner. Photogrammetric Engineering and Remote Sensing, 68(9), 925-932.

Persson, H., Wallerman, J., Olsson, H., & Fransson, J. E. (2013). Estimating forest biomass and height using optical stereo satellite data and a DTM from laser scanning data. Canadian Journal of Remote Sensing, 39(3), 251-262.

Peuhkurinen, J., Maltamo, M., Malinen, J., Pitkänen, J., & Packalén, P. (2007). Preharvest measurement of marked stands using airborne laser scanning. Forest Science, 53(6), 653-661.

Phillips, O., Baker, T., Feldpausch, T., & Brienen, R. (2009). RAINFOR field manual for plot establishment and remeasurement. Moore Foundation, Leeds, UK.

Pierce, K. B., Ohmann, J. L., Wimberly, M. C., Gregory, M. J., & Fried, J. S. (2009). Mapping wildland fuels and forest structure for land management: a comparison of nearest neighbor imputation and other methods. Canadian Journal of Forest Research, 39(10), 1901-1916.

Popescu, S. C. (2007). Estimating biomass of individual pine trees using airborne lidar. Biomass & Bioenergy, 31(9), 646-655. doi:10.1016/j.biombioe.2007.06.022

Popescu, S. C., & Wynne, R. H. (2004). Seeing the trees in the forest: Using lidar and multispectral data fusion with local filtering and variable window size for estimating tree height. Photogrammetric Engineering and Remote Sensing, 70(5), 589-604.

Popescu, S. C., & Zhao, K. (2008). A voxel-based lidar method for estimating crown base height for deciduous and pine trees. Remote Sensing of Environment, 112(3), 767-781. doi:10.1016/j.rse.2007.06.011

Popescu, S. C., Wynne, R. H., & Nelson, R. F. (2003). Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. Canadian Journal of Remote Sensing, 29(5), 564-577.

Popescu, S. C., Zhao, K. G., Neuenschwander, A., & Lin, C. S. (2011). Satellite lidar vs. small footprint airborne lidar: Comparing the accuracy of aboveground biomass estimates and forest structure metrics at footprint level. Remote Sensing of Environment, 115(11), 2786-2797. doi:10.1016/j.rse.2011.01.026

Rahman, M., & Gorte, B. (2009). Tree crown delineation from high resolution airborne lidar based on densities of high points. Paper presented at the Proceedings ISPRS Workshop Laserscanning 2009, September 1-2, France, IAPRS, XXXVIII (3/W8), 2009.

Reitberger, J., Schnorr, C., Krzystek, P., & Stilla, U. (2009). 3D segmentation of single trees exploiting full waveform LIDAR data. ISPRS Journal of Photogrammetry and Remote Sensing, 64(6), 561-574. doi:10.1016/j.isprsjprs.2009.04.002

Reutebuch, S. E., Andersen, H. E., & McGaughey, R. J. (2005). Light detection and ranging (LIDAR): An emerging tool for multiple resource inventory. Journal of Forestry, 103(6), 286-292.

Roberts, E. G., & Ross, R. D. (1965). Crown area of tree-growing loblolly pine and its apparent independence of age and site. Journal of Forestry, 43(3), 243-246.

Salas, C., Ene, L., Gregoire, T. G., Naesset, E., & Gobakken, T. (2010). Modelling tree diameter from airborne laser scanning derived variables: A comparison of spatial statistical models. Remote Sensing of Environment, 114(6), 1277-1285. doi:10.1016/j.rse.2010.01.020

Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). Biometrika, 52, 591-&.

Shin, J., Temesgen, H., Strunk, J. L., & Hilker, T. (2016). Comparing Modeling Methods for Predicting Forest Attributes Using LiDAR Metrics and Ground Measurements. Canadian Journal of Remote Sensing, 42(6), 739-765. doi:10.1080/07038992.2016.1252908

Silva, C. A., Crookston, N. L., Hudak, A. T., & Vierling, L. A. (2015). rLiDAR: An R package for reading, processing and visualizing LiDAR (Light Detection and Ranging) data, version 0.1. Retrieved from https://cran.r-project.org/web/packages/rLiDAR/index.html

Silva, C. A., Hudak, A. T., Vierling, L. A., Loudermilk, E. L., O'Brien, J. J., Hiers, J. K., Khosravipour, A. (2016). Imputation of Individual Longleaf Pine (Pinus

palustris Mill.) Tree Attributes from Field and LiDAR Data. Canadian Journal of Remote Sensing, 42(5), 554-573. doi:10.1080/07038992.2016.1196582

Silva, C. A., Klauberg, C., Hudak, A. T., Vierling, L. A., Liesenberg, V., Carvalho, S. P. C. E., & Rodriguez, L. C. E. (2016). A principal component approach for predicting the stem volume in Eucalyptus plantations in Brazil using airborne LiDAR data. Forestry, 89(4), 422-433. doi:10.1093/forestry/cpw016

Sironen, S., Kangas, A., & Maltamo, M. (2010). Comparison of different non-parametric growth imputation methods in the presence of correlated observations. Forestry, 83(1), 39-51. doi:10.1093/forestry/cpp030

Soille, P. (1999). Morphological phase unwrapping. Optics and Lasers in Engineering, 32(4), 339-352.

Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. In *Australian conference on artificial intelligence* (Vol. 4304, pp. 1015-1021).

Stage, A. R., & Crookston, N. L. (2007). Partitioning error components for accuracy-assessment of near-neighbor methods of imputation. Forest Science, 53(1), 62-72.

Steininger, M. K. (2000). Satellite estimation of tropical secondary forest above-ground biomass: data from Brazil and Bolivia. International Journal of Remote Sensing, 21(6-7), 1139-1157.

Stephens, P. R., Kimberley, M. O., Beets, P. N., Paul, T. S. H., Searles, N., Bell, A., Broadley, J. (2012). Airborne scanning LiDAR in a double sampling forest carbon inventory. Remote Sensing of Environment, 117, 348-357. doi:10.1016/j.rse.2011.10.009

Stibig, H. J., & Malingreau, J. P. (2003). Forest cover of insular Southeast Asia mapped from recent satellite images of coarse spatial resolution. Ambio, 32(7), 469-475.

St-Onge, B., Hu, Y., & Vega, C. (2008). Mapping the height and above-ground biomass of a mixed forest using lidar and stereo Ikonos images. International Journal of Remote Sensing, 29(5), 1277-1294. doi:10.1080/01431160701736505

Suarez, J. C., Ontiveros, C., Smith, S., & Snape, S. (2005). Use of airborne LiDAR and aerial photography in the estimation of individual tree heights in forestry. Computers & Geosciences, 31(2), 253-262. doi:10.1016/j.cageo.2004.09.015

Sumida, A., Miyaura, T., & Torii, H. (2013). Relationships of tree height and diameter at breast height revisited: analyses of stem growth using 20-year data of an even-aged Chamaecyparis obtusa stand. Tree physiology, 33(1), 106-118.

Symington, C. F. (1943). Foresters' manual of dipterocarps: U. of Malaya.

Team, R. D. C. (2016). R: a language and environment for statistical computing. Vienna, Austria. Retrieved from http://www.R-project.org

Temesgen, H., & Ver Hoef, J. M. (2015). Evaluation of the spatial linear model, random forest and gradient nearest-neighbour methods for imputing potential productivity and biomass of the Pacific Northwest forests. Forestry: An International Journal of Forest Research, 88(1), 131-142.

Temesgen, H., Barrett, T. M., & Latta, G. (2008). Estimating cavity tree abundance using nearest neighbor imputation methods for western Oregon and Washington forests.

Thapa, R. B., Watanabe, M., Motohka, T., Shiraishi, T., & Shimada, M. (2015). Calibration of aboveground forest carbon stock models for major tropical forests in central Sumatra using airborne LiDAR and field measurement data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8(2), 661-673.

Thomas, V., Treitz, P., McCaughey, J. H., & Morrison, I. (2006). Mapping stand-level forest biophysical variables for a mixedwood boreal forest using lidar: an examination of scanning density. Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere, 36(1), 34-47. doi:10.1139/X05-230

Tomppo, E., Malimbwi, R., Katila, M., Mäkisara, K., Henttonen, H. M., Chamuya, N., . . . Otieno, J. (2014). A sampling design for a large area forest inventory: case Tanzania. Canadian Journal of Forest Research, 44(8), 931-948.

Tonolli, S., Dalponte, M., Neteler, M., Rodeghiero, M., Vescovo, L., & Gianelle, D. (2011). Fusion of airborne LiDAR and satellite multispectral data for the estimation of timber volume in the Southern Alps. Remote Sensing of Environment, 115(10), 2486-2498.

Toutin, T., & Gray, L. (2000). State-of-the-art of elevation extraction from satellite SAR data. ISPRS Journal of Photogrammetry and Remote Sensing, 55(1), 13-33. doi:Doi 10.1016/S0924-2716(99)00039-8

Tsui, O. W., Coops, N. C., Wulder, M. A., Marshall, P. L., & McCardle, A. (2012). Using multi-frequency radar and discrete-return LiDAR measurements to estimate above-ground biomass and biomass components in a coastal temperate forest. ISPRS Journal of Photogrammetry and Remote Sensing, 69, 121-133. doi:10.1016/j.isprsjprs.2012.02.009

UNFCCC. (2006). Climate Change: Impacts, Vulnerabilities and Adaptation in Developing Countries. Bonn, Germany

UNFCCC. (2007). The United Nations Framework Convention in Climate Change Conference. Bali, Indonesia.

UNFCCC. (2010). The Cancun Agreements: Outcome of the work of the Ad Hoc Working Group on Long-term Cooperative Action under the Convention. Cancun, Mexico

UNFCCC. (2011). Report of the Global Environment Facility to the Seventeenth Session of the Parties to the United Nations Framework Convention on Climate Chane. Durban, South Africa.

van Aardt, J. A. N., Wynne, R. H., & Oderwald, R. G. (2006). Forest volume and biomass estimation using small-footprint lidar-distributional parameters on a per-segment basis. Forest Science, 52(6), 636-649.

Vastaranta, M., Holopainen, M., Haapanen, R., Yu, X., Melkas, T., Hyyppä, J., & Hyyppä, H. (2009). Comparison between an area-based and individual tree detection method for low-pulse density ALS-based forest inventory. Proceedings of Laser Scanning, Paris, France, 1-2.

Vastaranta, M., Holopainen, M., Yu, X., Haapanen, R., Melkas, T., Hyyppä, J., & Hyyppä, H. (2011). Individual tree detection and area-based approach in retrieval of forest inventory characteristics from low-pulse airborne laser scanning data. Photogrammetric Journal of Finland, 22(2), 1-13.

Vauhkonen, J., Korpela, I., Maltamo, M., & Tokola, T. (2010). Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. Remote Sensing of Environment, 114(6), 1263-1276.

Vazirabad, Y. F., & Karslioglu, M. O. (2011). Lidar for Biomass Estimation. Biomass - Detection, Production and Usage, 3-26.

Venables, W. N. & Ripley, B. D. (2002). Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

Ver Hoef, J. M., & Temesgen, H. (2013). A comparison of the spatial linear model to nearest neighbor (k-NN) methods for forestry applications. PloS one, 8(3), e59129.

Vincent, G., Sabatier, D., Blanc, L., Chave, J., Weissenbacher, E., Pelissier, R., . . . Couteron, P. (2012). Accuracy of small footprint airborne LiDAR in its predictions of tropical moist forest stand structure. Remote Sensing of Environment, 125, 23-33. doi:10.1016/j.rse.2012.06.019

Vincent, L., & Soille, P. (1991). Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Transactions on Pattern Analysis & Machine Intelligence (6), 583-598.

Vo, A.-V., Truong-Hong, L., Laefer, D. F., & Bertolotto, M. (2015). Octree-based region growing for point cloud segmentation. ISPRS Journal of Photogrammetry and Remote Sensing, 104, 88-100.

Wan-A-Kadir, W. R., & Vijayanathan, J. (2015). FRIM Tropical Man-Made Forest Research Park.

Wang, L. (2010). A Multi-scale Approach for Delineating Individual Tree Crowns with Very High Resolution Imagery. Photogrammetric Engineering and Remote Sensing, 76(4), 371-378.

Wang, Q., Ni, J., & Tenhunen, J. (2005). Application of a geographically-weighted regression analysis to estimate net primary production of Chinese forest ecosystems. Global Ecology and Biogeography, 14(4), 379-393.

Wang, X., Ouyang, S., Sun, O. J., & Fang, J. (2013). Forest biomass patterns across northeast China are strongly shaped by forest height. Forest Ecology and Management, 293, 149-160. doi:10.1016/j.foreco.2013.01.001

Wang, Y. S., Weinacker, H., & Koch, B. (2008). A Lidar point cloud based procedure for vertical canopy structure analysis and 3D single tree modelling in forest. Sensors, 8(6), 3938-3951. doi:10.3390/s8063938

Wan-Mohd-Jaafar, W., Woodhouse, I., Silva, C., Omar, H., & Hudak, A. (2017). Modelling Individual Tree Aboveground Biomass Using Discrete Return LiDAR in Lowland Dipterocarp Forest of Malaysia. Journal of Tropical Forest Science, 465-484.

White, J. C., Wulder, M. A., Varhola, A., Vastaranta, M., Coops, N. C., Cook, B. D., Woods, M. (2013). A best practices guide for generating forest inventory attributes from airborne laser scanning data using an area-based approach. Forestry Chronicle, 89(6), 722-723. doi:10.5558/tfc2013-132

White, J. C., Coops, N. C., Wulder, M. A., Vastaranta, M., Hilker, T., & Tompalski, P. (2016). Remote sensing technologies for enhancing forest inventories: A review. *Canadian Journal of Remote Sensing*, *42*(5), 619-641.

Wile, B. C. (1964). Crown size and stem diameter in red spruce and balsam fir. Department of Forestry Publication (1056), 1-9.

Williams, M., Ryan, C. M., Rees, R. M., Sarnbane, E., Femando, J., & Grace, J. (2008). Carbon sequestration and biodiversity of re-growing miombo woodlands in Mozambique. Forest Ecology and Management, 254(2), 145-155. doi:10.1016/j.foreco.2007.07.033

Wolf, B. M., & Heipke, C. (2007). Automatic extraction and delineation of single trees from remote sensing data. Machine Vision and Applications, 18(5), 317-330. doi:10.1007/s00138-006-0064-9.

Woods, M., Lim, K., & Treitz, P. (2008). Predicting forest stand variables from LIDAR data in the Great Lakes St. Lawrence Forest of Ontario. *The Forestry Chronicle*, *84*(6), 827-839.

Woods, M., Pitt, D., Penner, M., Lim, K., Nesbitt, D., Etheridge, D., & Treitz, P. (2011). Operational implementation of a LiDAR inventory in Boreal Ontario. *The Forestry Chronicle*, *87*(4), 512-528.

Wulder, M. A., White, J. C., Nelson, R. F., Naesset, E., Orka, H. O., Coops, N. C., Gobakken, T. (2012). Lidar sampling for large-area forest characterization: A review. Remote Sensing of Environment, 121, 196-209. doi:10.1016/j.rse.2012.02.001.

Wulder, M. A., White, J. C., Stinson, G., Hilker, T., Kurz, W. A., Coops, N. C., Trofymow, J. A. (2010). Implications of differing input data sources and approaches upon forest carbon stock estimation. Environ Monit Assess, 166(1-4), 543-561. doi:10.1007/s10661-009-1022-6.

Wulder, M. A., Bater, C. W., Coops, N. C., Hilker, T., & White, J. C. (2008). The role of LiDAR in sustainable forest management. The Forestry Chronicle, 84(6), 807-826.

Wulder, M., Niemann, K. O., & Goodenough, D. G. (2000). Local maximum filtering for the extraction of tree locations and basal area from high spatial resolution

imagery. Remote Sensing of Environment, 73(1), 103-114. doi:Doi 10.1016/S0034-4257(00)00101-2.

Wyatt-Smith, J. (1964). A preliminary vegetation map of Malaya with descriptions of the vegetation types. Journal of Tropical Geography, 18, 200-213.

Yu, X., Hyyppä, J., Litkey, P., Kaartinen, H., Vastaranta, M., Holopainen, M. (2017). Single-Sensor Solution to Tree Species Classification Using Multispectral Airborne Laser Scanning. Remote Sensing, 9(2), 108.

Zhao, F., Guo, Q., & Kelly, M. (2012). Allometric equation choice impacts lidar-based forest biomass estimates: A case study from the Sierra National Forest, CA. Agricultural and forest meteorology, 165, 64-72.

Zhao, K. G., Popescu, S., & Nelson, R. (2009). Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. Remote Sensing of Environment, 113(1), 182-196. doi:10.1016/j.rse.2008.09.009

Zhao, K. G., Popescu, S., Meng, X. L., Pang, Y., & Agca, M. (2011). Characterizing forest canopy structure with lidar composite metrics and machine learning. Remote Sensing of Environment, 115(8), 1978-1996. doi:10.1016/j.rse.2011.04.001

Zhen, Z., Quackenbush, L. J., & Zhang, L. J. (2014). Impact of Tree-Oriented Growth Order in Marker-Controlled Region Growing for Individual Tree Crown Delineation Using Airborne Laser Scanner (ALS) Data. Remote Sensing, 6(1), 555-579. doi:10.3390/rs6010555

Zhen, Z., Quackenbush, L. J., & Zhang, L. J. (2016). Trends in Automatic Individual Tree Crown Detection and Delineation-Evolution of LiDAR Data. Remote Sensing, 8(4). doi:ARTN 333

Zhen, Z., Quackenbush, L. J., Stehman, S. V., & Zhang, L. J. (2015). Agent-based region growing for individual tree crown delineation from airborne laser scanning (ALS) data. International Journal of Remote Sensing, 36(7), 1965-1993. doi:10.1080/01431161.2015.1030043

Zolkos, S. G., Goetz, S. J., & Dubayah, R. (2013). A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. Remote Sensing of Environment, 128, 289-298. doi:10.1016/j.rse.2012.10.017.